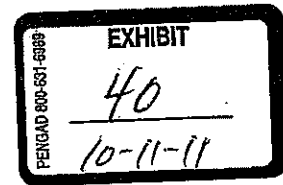


# **Exhibit 14**

UNITED STATES DISTRICT COURT  
FOR THE DISTRICT OF NEW JERSEY



In re: Aetna UCR litigation,

MDL NO. 2020

This Document Relates To: ALL CASES

MASTER FILE NO. 2:07-CV-3541 (FSH)

DECLARATION OF STEPHEN FOREMAN, PHD, JD, MBA

**Introduction**

1. I have reviewed reports and materials submitted by defendants' experts critiquing my Merits Report of August 9, 2010 ("the August 9 Report" or "the Report"). The reports cite numerous examples of "error" in the Report.
2. What many of what the Defendants' experts claim to be errors are not errors but attempts to bias, alter and otherwise obscure findings and results.
3. The Report was produced under pressure of limited time, limited resources and limited knowledge. I stated in the Report that, to the extent, that additional information and knowledge became available that corrections and changes to the Report would be made.
4. There are some errors in the Report which I have attempted to correct. None of them provides a material change in the Report or my conclusions.
5. Defendants' experts complain about a range of decisions underlying the evaluation of downward bias in the Ingenix data such as the use of billed claim lines rather than "units" in evaluating contributor data percentiles, the selection of the most common 300 CPT codes and the most common 350 CPT codes for percentile comparison, the attempt to provide a comparison of Ingenix product percentiles that would have been used to adjudicate contributor claims (timing issue) and the elimination of comparisons where the lack of data suggested that percentiles are essentially random. The basic approach in the Report was and remains sound.
6. In order to demonstrate the immaterial nature of these complaints and to extend and expand our understanding of issues with the Ingenix products we undertook additional investigation in the form of

- evaluation of the influence of considering units on percentile comparison results.
  - a comparison of contributor data percentiles and Ingenix product percentiles for 500 CPT codes in all 421 Ingenix geozips (94% of all contributor claims) for 2006 through 2008 (210,500 comparisons each year) that includes an analysis of the impact of including and excluding comparisons where there are few observations.
  - a comparison of contributor data percentiles and Ingenix product percentiles for all (5137) CPT codes used in the Ingenix product data in all 421 Ingenix geozip for 2006 through 2008 (2.2 million comparisons each year). This comparison factored differences in units into the analysis, evaluated all of the data as a subset, provided percentile comparisons for contemporaneous Ingenix products as well as prior period Products and included CPT/geozip combinations of any size.
7. The 500 and the 5000 CPT studies verify and confirm the findings of the 300 and the 350 CPT studies contained in the Report – that Ingenix product percentiles are biased downward and that the downward bias for medical and surgical CPTs substantially exceeds the 11.2% amount used for damage calculation.

#### **Background**

8. Ingenix collects billed charge data from a limited set of health insurers and employee benefit plans, processes that data using procedures explained in a 600-page manual, produces statistics in the form of percentiles using the data, and sells the percentile product to health insurers (many of which are the same health insurers that contribute to the data) that include Aetna.
9. Health insurers (including Aetna) use the percentiles to limit amount they reimburse (paid) to patients and physicians for out of network medical care under contracts that specify that they will pay the lesser of billed charge or "usual, customary and reasonable" (UCR) amounts (sometimes called "reasonable and customary" or R&C amounts).
10. UCR has been used for decades by Medicare and health insurers. It is generally understood as a way to provide an upper bound on reimbursement based on "comparable" or comparative medical care prices.
11. As applied in practice the concept of limiting billed charges based on UCR includes three commonly understood concepts:
- billed charges for the "same or similar services"
  - billed charges in the "community"
  - billed charges that are "contemporaneous"
12. Ingenix and the health insurers construct and apply two products- MDR and PHCS-to limit

patient and physician reimbursement.<sup>1</sup> In so doing, they warrant (in order to comply with their contracts) that the application of the percentile data limits reimbursement because the percentile data indicates contemporaneous usual, customary and reasonable values for the same or similar services in the community. Having developed and applied the percentiles for limiting reimbursement it is incumbent on Ingenix and the health insurers to assure that the percentiles themselves are accurate (unbiased) and represent contemporaneous comparisons for the same or similar services in the medical care markets where they are used. Having elected to use the UCR limitations, Ingenix and the health insurers assume a responsibility for accuracy, currency and relevance in their use of the percentile data.

13. The reports on the Merits by Dr. Siskin and Dr. Foreman identified a number of potential concerns with the Ingenix products including:
  - arbitrary elimination of data in a process that Ingenix has called the "high low screen"
  - collection of data from substantially less than all health insurers with no attempt to discern whether the data collected are representative of billed charges for the same or similar services in a medical care market
  - lack of contemporaneousness in the PHCS percentile data
  - use of broad geographic areas (identified by the first three digits of zip codes) as a surrogate for communities or medical care markets
  - failure to consider provider qualifications and training (specialty)
  - use of insufficient data to provide appropriate levels of confidence in the percentile values reported
  - use of "derived" data to generate and report percentile values which bear no relationship to contemporaneous medical care services in the market
14. Rather than providing scientific analysis of the accuracy of the Ingenix percentiles, the reports of defendants' experts offer voluminous erroneous, immaterial and irrelevant commentary. Substantial and important parts of the reports by defendants' experts provide support for the conclusions that the Ingenix product percentiles are biased, are systematically biased downward in a way that benefits health insurers, have been arbitrarily adopted for convenience and bear no reasonable relationship to UCR or R&C.
15. Defendants' experts critique numerous aspects of the Reports on the Merits in a cacophony of confusing and contradictory arguments but generally fail to offer any analysis that shows that Ingenix percentiles meet the concept of UCR or R&C.

---

<sup>1</sup> To the extent that they have used Ingenix products defendant health insurers primarily have used PHCS.

**Professor Slottje's critique**

**The nature of UCR**

16. In the introduction to Section III of his report Dr. Slottje ellipses portions of responses to complicated and convoluted questions, including hypotheticals, to paint a fully confusing description of my understanding of UCR and how it relates to the Ingenix percentiles and statistical concepts. "Because Dr. Foreman cannot define with statistical concepts what he means when he uses the term UCR throughout his report he is contradicting himself when he renders opinions that techniques used by Ingenix in creating the PHCS database would not establish UCR because they fail to comply with various statistical concepts."
17. Dr. Slottje's ellipses and obfuscation go directly to the essence of his (and defendants') justification for the process of collecting data from certain – but not all – health insurers, the manipulation of the data, the production of what Ingenix calls percentiles from the data and the use of the so-called percentiles to limit how patients and providers are reimbursed. The justification is that UCR is whatever Dr. Slottje or the health insurers determine it to be. If the data collected from health insurers over-counts large health insurers with lower billed charges (as the contributor data analysis in the Report shows)- no problem. Arbitrary elimination of data to make the percentiles any value – no problem. Use of old data to build percentiles – OK. Reimbursement of a brain surgeon at the same rate as a chiropractor – so what? Concluding that physicians in rural West Virginia practice in the same market as physicians in downtown Pittsburgh – or downtown Manhattan for that matter – why not? If we need to get rid of data or we need to merge areas or ignore specialization and training or medical cost inflation – anything goes.
18. To be absolutely clear - it is my expert opinion based on 30-years' experience as a hospital administrator, university professor, health care consultant and as a consumer who has lived through the implementation and application of UCR that:
  - The term UCR has meaning that is understood throughout the health care industry – and is understood by the public and the health insurance industry.
  - When health insurers hold themselves out as paying based on UCR it is everyone's belief (patients, providers, employers, regulators) that they will pay in a manner consistent with the industry's and the public's understanding of the term UCR.
  - If health insurers are legally able, consistent with their UCR contracts (a legal opinion for determination by the court), to limit provider payment using billed charge "percentiles" they have determined to make payment using a concept that is at least in part statistical.

- Patients, providers, employers and government officials do not understand the term “percentile” nor do they understand how percentiles are calculated, but they have a right to expect that the percentile has statistical meaning and that the percentiles are being developed in accordance with sound scientific principles.
- To the extent that the percentiles used to payment or reimbursement are developed and applied in ways that are arbitrary or biased then the term percentile has no meaning and their use for UCR has no meaning and the perception of patients, providers, employers and regulators regarding payment is undermined.
- Because they hold and use the data, the producer of the percentiles and the health insurers have the burden to establish the scientific accuracy of the process from data collection to production of the percentiles to use in reimbursement.

#### **Data elimination**

19. Dr. Slottje questions Dr. Siskin's critique of the Ingenix practice of automatically eliminating data when it exceeds a specified high value or is less than a specified low value on the basis that "an outlier must be incorrect or invalid to warrant its removal." Dr. Slottje states that this is merely Dr. Siskin's opinion which is not supported by statistical theory. He contends that there is an extensive literature on outliers and that the choice of outlier detection methodology requires value judgments as does the decision of whether or not to remove an outlier.<sup>2</sup>
20. In essence, Dr. Slottje's position is that it is appropriate and scientifically justified to automatically eliminate 6% to 7% of the data (what Dr. Slottje found to be eliminated with the high low screen) used when calculating percentiles because they are outliers – and that outlier identification based on prior period – not current period values – is appropriate and scientifically justified. There is no support for this in any scientific literature. To the contrary, scientific principles condemn it and his own empirical work rebuts it.
21. Dr. Slottje assumes without question that the Ingenix data removal process is an appropriate outlier detection and removal procedure. This is clearly wrong for a substantial number of reasons.
22. First, the Ingenix process does not detect and eliminate outliers. It is an arbitrary irrelevant process that eliminates a large amount of data in a way that (based on Dr. Slottje's own studies) provides downward bias to the percentiles in the PHCS and MDR products.
23. The Ingenix data removal process is automatic. It does not involve any value judgment

---

<sup>2</sup> Rebuttal Report of Dr. Daniel J Slottje to the August 9, 2010 Expert Report of Dr. Foreman and Dr. Siskin (Slottje) at pages 5-6 and later at pages 9-10.

whatsoever. The process uses "inter percentile ranges"<sup>3</sup> that are based on scientific literature and "gate" multipliers that are not based on scientific literature and bear no relationship to data distributions. If the Ingenix data removal process is to be justified as outlier removal-which it is not- it is subject to the same criticism that Dr. Slottje offers for Dr. Siskin.

24. Dr. Slottje is correct that there is a large literature on outlier identification and removal. None of it supports his observations or justifies the Ingenix processes. The literature identifies the outlier problem as instances where extreme values bias statistics. That statistics subject to such bias include mean values and linear regression model coefficients.<sup>4</sup> The Ingenix products are not based on mean values or regression model coefficients.
25. In general, outliers have been described using multiples of interquartile ranges (Tukey) using "weak" (1.5) and "strong" (3) multipliers applied to the first and third quartile values. A general rule of thumb provides for identification of outliers as more than three standard deviations from the mean in a normal distribution. Thus less than 0.3% of a normal distribution would be expected to be outliers. The Ingenix process identifies 6% to 7% of the data contributed to it (of a distribution that is clearly not normal) as "outliers" using a process that has not been developed in peer reviewed scientific literature. As such, the identification process is not aimed at identifying outliers but arbitrarily eliminates data.
26. Second, there are statistics such as the median (and percentiles) that are robust to outliers. This has led to the development of the field of "robust statistics"-a literature that applies ideas about robustness to avoid elimination of outliers from data.<sup>5</sup> Because the percentile values are robust to outliers there is absolutely no need to eliminate outliers from the data used to construct percentile values. To the extent that data are arbitrarily eliminated in a way that biases percentiles the so-called outlier identification process will not only be unjustified but will provide affirmative distortions to the percentile data products.
27. Third, the Ingenix "gate" or "screen" values are not based on current or contemporaneous data but are developed from prior period data submissions. As a result, in an era of billed charge inflation (which as noted in the report on the merits is the case based on the Ingenix contributor data submissions) the gate or screen values automatically are biased downward.
28. Fourth, even though he fails to consider the lag effect in the Ingenix screen, Dr. Slottje's own analysis establishes and proves its downward bias.

---

<sup>3</sup> The 50th and 80th percentiles

<sup>4</sup> S. Morgenthaler, A Survey of Robust Statistics, Statistical Methods and Applications, Volume 15, pp 271-93 (2007), page 272.

<sup>5</sup> *Id.*

**Reverse engineering the data elimination**

29. Professor Slottje states that he used contributor data to reverse engineer the high low screen – by “putting the eliminated data back” to see what would happen. Professor Slottje found that “for roughly 85% to 88% of all CPT – geozip combinations based on actual charge data [740,000 combinations], Ingenix’s application of the “Hi/Lo scrub” had *zero* impact on the percentile amounts...” Adding back in the high and low charges that Ingenix excluded due to the “Hi/Lo scrub,” *lowered* the percentile amounts for 3% to 6% of all CPT geozip combinations.
30. Thus, Professor Slottje says: *“adding back in the high and low charges that Ingenix excluded due to the “Hi/Lo scrub” either lowered, or had no impact on the percentile amounts for 89%-94% of all CPT-geozip combinations.”*
31. This is true, but what about the times that adding back the high and low charges increased percentile amounts. What about the times that the high low screen reduced percentile values? In fact, from Professor Slottje’s own data the high-low screen reduced percentile values six percent to 12% of the time so that the ratio of percentile reductions far outweighs the percentile increases. And, this is a qualitative, not a quantitative, analysis.
32. Professor Slottje shows- without explaining why - that only about half of the contributor data received are “eligible” for inclusion in the data. Using the Ingenix term, half of the data are scrubbed out of the data set even before the high low-screen? After the high low screen only one third to one half of the data remains. This suggests that there are substantial problems with the data and that there are other issues that Professor Slottje has not discussed. Professor Slottje summarily dismisses these issues as “independent of the scrubbing allegations proffered by plaintiffs.”
33. If the Ingenix percentile data purport to represent the “population” of billed charges for the purpose of adjudicating UCR (which they clearly are not), the elimination of half of the data before the percentile values are calculated raises serious questions about whether the percentile data can or should ever be used for UCR.<sup>6</sup>
34. Section III of the Slottje Report states that the study is an “analysis of 2006-2008 contributor data. However, Table provides information only for 2007 and 2008. Also, on page 10 the report states “I also measured the volume of 2006-2008 contributor data that corresponds to CPT-geozip combinations not affected by the high low scrub.” Despite this, the analysis covers only two years of data, evaluating impact for 2007 and 2008 product releases.
35. Table 3 summarizes the findings. The summaries in Table 3 are at odds with the discussion that

---

<sup>6</sup> Ingenix has a long list of data cleaning procedures. Ingenix00001283. The scientific validity of these procedures has not been presented.



- precedes it. In the discussion Professor Slottje states that adding back in the high low charges lowered the percentile amounts for 3% to 6% of all combinations and (by implication) raised the percentile values 6% to 12% of the time. Table 3 states that with the high low screen percentile values were lower 3% to 11% of the time so that they were higher 6% to 11 percent of the time.
36. Most important, what the qualitative data analysis clearly shows is that the high low screen reduced percentile values 6% to 12% of the time compared with an increase 3% to 6% of the time - a ratio of 2:1 - 4:1 - fully consistent with the 300 CPT study and the 350 CPT study that Dr. Slottje criticizes and consistent with the 500 CPT and 5000 CPT studies described below. In other words, Professor Slottje's own analysis shows that the high low screen reduced percentile values two to four times more often than it is increased them - and his study fails to deal with any timing issues.
  37. Professor Slottje focuses on the values that stayed the same. The real impact on reimbursement as a result of the high-low screen is not the values that stayed the same but what happened to patients and providers who saw their reimbursement inexplicably drop or rise?<sup>7</sup>
  38. In Table 4 Professor Slottje calculates the dollar value and volume of charges *unaffected* by the high low screen. He says that this confirms that the high low scrub does not impact the vast majority of CPT geozip combinations and does not impact the vast majority of volume of charges or dollar amount of charges.
  39. Professor Slottje's conclusion is not dispositive. Why has he not calculated or reported the dollar value and volume of the instances where the high-low screen diminished percentiles and those when it did?
  40. It is possible that the 6% to 12% of percentile values that were decreased far outweigh the dollar value of the percentile values that increased. Such a finding would be dispositive on the issue of downward bias and would fully contradict the conclusions that Dr. Slottje draws from his data.
  41. Moreover, to say that because the high-low screen "only" had an adverse impact on 6 to 12% of CPT/geozip combinations because some proportion of claims were unaffected is insensitive to those who were actually injured. The level of adverse impact, 6% - 12%, applied to billions of claims per year representing hundreds of billions of dollars (shown by Dr. Slottje's own empirical analysis) affected millions of claims and billions of dollars in reimbursement.
  42. Professor Slottje's analysis recomputed percentile amounts for all "actual" CPT/geozip combinations for the data releases, between 740,000 and 800,000 of them. What the analysis fails to consider is that Ingenix products report percentile values when there are as few as nine.

---

<sup>7</sup> Actually, the class consists of claims where reimbursement was denied using the percentile data so those whose reimbursement was greater than it should have been are not members of the class.

As described below, statistical power analysis suggests that what is represented as the 80th percentile when there are nine claims is essentially random. Frequency analysis of MDR products that use the same underlying data as PHCS, found 529,000 of 720,000 CPT/geozip combinations had between nine and 254 observations -- inadequate to report percentile values with confidence.<sup>8</sup>

43. Put another way, if CPT/ geozip combinations with small numbers of claims (between nine and 254) are essentially random then at the low end a high low screen would be expected to raise some percentile values, leave them the same (a large portion of the time) a decrease them in other instances. The increases and the decreases should be in the same proportion. Thus, one might expect a high low screen like Ingenix' to have two separate statistical effects: the first to reduce high end expensive billed charges and the second to reduce stochasticism in the CPT/geozip combinations with very few claims. If half of the effect relates to the first phenomenon and a half to the second, we would expect that the screen would produce percentile value reductions 75% of the time and percentile value increases 25% of the time. This is exactly what Professor Slottje's empirical analysis shows.<sup>9</sup>
44. Professor Slottje fails to report why he measured the percentiles that he did. At the 75<sup>th</sup> percentile the high low screen would be expected to have less impact than at the 80<sup>th</sup> or the 90<sup>th</sup> (more commonly used by health insurers) since the high low screen is aimed at the highest values and disproportionately impacts higher level percentiles (80th, 85th, 90th and 95th).
45. As noted by Professor Siskin, censoring of data at the high and low ends theoretically provides a much greater dollar reduction at the high end than a commensurate the dollar addition at the low end.<sup>10</sup> Eliminating data relating to expensive procedures is not offset by eliminating a like amount of data for low cost procedures. If the same number of claims is eliminated at the top and the bottom of the distribution there will be a new downward bias in terms of total dollar value. If the number of high value charges eliminated is greater than the number of low value charges eliminated, the dollar value difference at the high end and the low end will produce a large reduction in percentile values and in reimbursement.
46. If, for example, the ratio of billed charge dollar amounts eliminated at the high end is ten times greater than the billed charge amounts eliminated at the low end, the four to one ratio of high end percentiles reduced to low end percentiles increased would translate into a ratio of 40:1 for the dollar value of the percentiles reduced compared to the low end values increased.

---

<sup>8</sup> *Id.*

<sup>9</sup> Moreover, for a great many of the "small cell" combinations all values will be the same so that a high low screen can be expected not to impact them.

<sup>10</sup> Franco v. CIGNA, Plaintiffs' Expert Report of Bernard R. Siskin, PhD., April 5, 2010.

47. In order to evaluate the implications of Professor Slottje's empirical results – that 3% - 6% of the time percentile values go up with the high low screen and 6%-12% of the time they go down, we used data from the 350 CPT Study that evaluated contributor data percentiles and PHCS percentiles for 350 of the most common procedures for all geozips - using 400 million claims from the first six months of 2008 – comparing them to the second PHCS release of 2007.
48. The qualitative results from extension of the 350 CPT Study were consistent with those reported by Professor Slottje: There were percentile values in PHCS that were lower than comparable contributor data percentiles, there were contributor values that were the same and there were contributor values that were higher.<sup>11</sup> At the 80<sup>th</sup> percentile the ratio of decreases to increases was about four to one.<sup>12</sup>
49. The relationship between the number of claims, the direction of increase or decrease and their dollar value is most informative. Percentile values for 645 million claims (counts as reported for the PHCS percentiles) were calculated for the contributor data. The average 80<sup>th</sup> percentile value was \$344 for claims whose values were less in the PHCS than in the contributor data and \$138 for claims whose value was greater in PHCS than in the contributor data – the value of the screen's decreases was more than double the increases.
50. As shown in Table 1, high billed charge values dropped and low billed charge values increased – but high value claims dropped four times as often as low value claims increased while the value of the high end declines was more than double the value of the low end increases. The total dollar impacts – a function of frequency and value – for reductions compared with increases was eleven to one.
51. Similarly, at the 90<sup>th</sup> percentile the value of CPT/geozip percentiles that were reduced from the contributor data percentiles to the product percentiles was \$434, while the value of the percentiles that increased was \$169.
52. For additional qualitative analysis we separated the CPT/ geozip combinations into those higher than the mean 80<sup>th</sup> percentile value and lower than the mean 80<sup>th</sup> percentile value. In the upper half of the distribution percentile values were lower in the PHCS than the contributor value 81% of the time, they were the same 10% of the time and they were higher nine percent of the time. In the lower half PHCS values were lower 62% of the time, the same 20% of the time and higher 20% of the time.
53. Both results are consistent with the hypothesis that the high low screen reduces high end values

---

<sup>11</sup> Professor Slottje found many more unchanged than in this study but he did not consider all of the challenged processes such as the impact of time.

<sup>12</sup> Foreman Expert Report, supra at p. 88.

and will reduce some and increase some percentile values when there are few claims to provide good confidence in the percentile value products.

54. If Professor Slottje had considered frequency and dollar value of the CPT / geozip combinations in his study that dropped due to the high low screen to those that increased, he would have found that the value of the reductions were eleven times greater than the increases.

**Dollar Value of Percentiles That Decreased from the Contributor Data to the PHCS Product  
Compared to Dollar Values of Percentiles That Increased**

Percentile	50	60	70	75	80	85	90	95
Equal values	131.20	138.90	146.68	150.06	152.86	160.28	175.15	434.37
Values reduced	229.47	254.69	289.53	314.19	344.46	381.55	434.37	522.30
Values increased	119.10	124.62	130.50	133.72	137.74	150.04	169.05	201.41

55. In short, Professor Slottje's critique of Dr. Siskin's observations about the Ingenix data elimination is erroneous based on scientific literature, practice and his own empirical work. It is not appropriate to eliminate large amounts of data when computing percentiles on the premise that the data constitute outliers that influence the percentile statistics and such elimination changes – biases the resulting percentiles both downward and upward with the quantity and dollar value of the downward bias far greater than the quantity and dollar value of upward bias.
56. Dr. Slottje also "extracts" discussion the August 9<sup>th</sup> Merits Reports and my two day deposition relating to the absence of scientific basis or rationale for elimination of the data.
57. Dr. Slottje states that "the scientific basis for both identifying and removing outliers is well established – it is based on the understanding that an outlier can distort inferences or cause other problems when one is attempting to draw conclusions about certain characteristics of a distribution or set of data citing a dated version of Barrett and Lewis.<sup>13</sup>
58. Dr. Slottje's entire analysis is based on outlier theory relating to statistics that are weakened or degraded by outliers – means or model coefficients – because they are "variance" sensitive. In order to motivate his particularly inapt observations and example, Dr. Slottje's discussion uses pointedly imprecise terms such as "certain characteristics of a distribution or set of data.
59. A question (the response to which is taken out of context) from my November 1st and 2d deposition was whether the inclusion of Bill Gates' income would distort the "level of inequality" in the income distribution for the town. The question knowingly used imprecise language in order to generate a response that could be attacked.
60. As expressed in my deposition testimony and in my Report, inclusion of Mr. Gates' income in computing the mean level of income in his town would distort *the mean* as a statistic relating to

<sup>13</sup> Slottje Rebuttal Report at page 15.

income level. However, the inclusion of Mr. Gates's income *would not distort the median* level of income -- nor would it distort the 60th percentile, the 70th percentile, the 75th percentile, or any other percentile of income. Indeed, to disregard, ignore or eliminate Mr. Gates income in the context of a discussion of income levels (as would Dr. Slottje) – simply because it distorts the mean – would be unfortunate and potentially wrong.

61. A simple example illustrates this point. Suppose Mr. Gates has a one million dollar salary and shares his town with only four others – with income levels of \$20,000, \$40,000, \$60,000 and \$80,000. The mean or average level of income in the town – including Mr. Gates is \$240,000. The mean doesn't say much accurately about either Mr. Gates (distorted low) or his fellow townspeople (distorted high) in terms of how wealthy people are in town. However, if income tax collections are at issue the mean income level would describe the town pretty well. If you throw out Mr. Gates income when estimating tax collections it would be a mistake. If income disparity is the question of interest it would be totally distorting to ignore Mr. Gates' income. The median level of income for this little town is \$60,000 with Mr. Gates included and \$40,000 if his income is excluded. Clearly, inclusion of Mr. Gates' income is proper if one is interested in the true median – exclusion of his income would improperly distort the median.<sup>14</sup>
62. To extend the concept to the UCR setting – not Bill Gates -suppose we have two small towns with five cardiac surgeries in each - payment for which is adjudicated using percentile data.
- In the first town physicians charge \$2000, \$3000, \$4000, \$8000 and \$10,000.
  - In the second town \$2000, \$3000, \$4000, \$8000 and \$100,000 where the \$100,000 charge relates to a university hospital's charge to out of country royalty for a very involved cardiac procedure – the charge is not a mistake.
  - UCR payment is limited to the 80<sup>th</sup> percentile.
  - In both towns – the 80<sup>th</sup> percentile limit is \$8,000.
  - If calculation of UCR in the second town “throws out” the \$100,000 charge the 80<sup>th</sup> percentile value in town two drops.
  - The physician in town one who charges \$8000 will be paid in full while the physician in town two who charges the same amount will be paid less – solely based on elimination of one high end billed charge.
  - The \$100,000 billed charge may or may not have been an outlier. It had no effect, however, on the percentile values in town two and it would be a mistake to ignore it when reimbursing UCR.

---

<sup>14</sup> Just as exclusion of outliers distorts percentile values in the Ingenix data.

63. Dr. Slottje is either unable or unwilling to understand or pointedly ignores the difference between statistics like the mean that are subject to distortion by outliers and statistics (importantly including percentiles) which are robust or impervious to outliers.
64. Dr. Slottje also quotes out of context (in discussion applicable to means not percentiles) that "modern exploratory data analysts generally agree that it is usually necessary to eliminate "a few" data points in order to describe the data by a reasonably simple model." The Ingenix percentile process provides percentiles, not data description using a model.
65. Dr. Slottje cites a number of examples (none of them from peer reviewed statistical journals or texts) that purport to justify automatic removal of outliers from large data sets. None of them justifies automatic elimination of outliers in percentile contexts.
66. Dr. Slottje misconstrues and distorts the discussion involving a Tukey screen in the August 9<sup>th</sup> Merits Report. He uses it as justification of automatic application of an outlier methodology in all circumstances. The Tukey screen example was not intended to be a perfect exploration of the Ingenix high low screen. It was intended to show – using data from Aetna and CIGNA – how a Tukey screen would bias data percentiles in general and particularly, it could bias percentiles downward.
67. Dr. Slottje purports to use "the same methodology" as his July 30<sup>th</sup> report to measure the impact of Ingenix' high low scrub. The methodology is confusing and distorting. Dr. Slottje's July 30<sup>th</sup> approach "adds back" eliminated data "actual impact of inclusion of charges otherwise removed" such that the signs in his tables on page 19 are the opposite of common sense (they constitute a double negative in a sense).
68. For the 300 CPT comparison the Ingenix data elimination reduced percentiles in the Ingenix products (adding back the eliminated data increased the percentiles) relative to the 300 CPT/300 geozip combinations 9% to 10% of the time compared to increases 6% of the time compared to increases for the 80<sup>th</sup> percentile and 90<sup>th</sup> percentiles and reduced the percentiles 12% to 14% of the time vs. increases 4% to 6% of the time. Dr. Slottje's analysis establishes that the elimination of Ingenix data – the hi/lo scrub - provided downward bias to CPT/geozip percentiles.
69. Dr. Slottje also distorts the analysis by confounding the high low screen analysis and the overall downward bias analysis. His high low screen study looks only at the impact of the data removal – not at the impact of a large number of processing factors that include ignoring the influence of time.

#### **Representativeness**

70. Dr. Slottje criticizes Dr. Siskin's observation that the data received from the contributors is a

convenience sample and is, therefore, likely to be biased. The principal thrust of the critique is that Dr. Siskin has not tested the Ingenix convenience sample for bias. Dr. Slottje criticizes Dr. Siskin for concluding that the convenience sample is biased without testing.

71. Dr. Slottje's approach is fully backward. Ingenix provides percentile data to health insurers in the form of MDR and PHCS. Their use is proper only if the percentiles are not biased. The null hypothesis in any test must be that the Ingenix percentiles-since they do not represent a population-are in fact biased. It is incumbent on Ingenix (and Dr. Slottje) to empirically establish the lack of bias in order to reject this null hypothesis. Both Ingenix and Dr. Slottje know how to conduct such studies but have failed to do so. Without empirical evidence to reject the null hypothesis (that the contributor data is biased) Dr. Slottje and Ingenix cannot conclude that the percentiles produced by the Ingenix contributor data are unbiased.
72. Dr. Slottje quotes with approval from my August 9 report on the merits "because the entire population of billed charges is not known there is no way to scientifically prove or disprove whether the Ingenix billed charges are representative of all billed charges. The hypothesis that the Ingenix percentiles represent UCR can neither be proven nor disproven." However, Dr. Slottje appears to use this as justification for continued use of the Ingenix percentiles for UCR determinations. Nothing could be further from the truth. Because Ingenix cannot establish the absence of bias in the contributor data Ingenix cannot claim that the percentiles produced by its products are free from bias and, accordingly, there is no scientific justification for concluding that the Ingenix percentiles are appropriate for use in UCR determinations.
73. Dr. Slottje uses the testimony of Dr. Siskin referring to "census" taken by Ms. Faddis when she collected information regarding provider charges in her area to attempt to justify how Ingenix collects data as a "census" thus making it a "population." "A population of data that Ingenix collects is the precise data they state they are collecting in their manual. Ingenix's data collection effort is a census because the data collected is the entire population of data. There is no sampling occurring. If no sampling is occurring, then no convenience sampling is occurring and that no sample bias exists.
74. Dr. Slottje's observations here are clearly distorting. His characterization confuses and obfuscates the difference between data collected from the contributors and contemporary billed charges for like or similar services in a community or medical market.
75. In a discussion or analysis that relates solely to data obtained from contributors, the data might be characterized or defined as the population of contributor data. However, where the discussion or analysis relates to the population of billed charges in a medical market and, as is clearly the case, the Ingenix contributor data constitute substantially less than all of the population of billed



- charges in the market, then the contributor data are a subset of the population – thereby a sample.
76. Solely from the standpoint of the percentiles constructed by Ingenix, those percentiles represent percentiles of the contributor data-and if all of the contributor data is used (which they are not)-the percentiles represent percentiles of the population of contributor data. When the percentiles are used to adjudicate UCR by health insurers (a process that both Ingenix and the health insurers clearly understand is occurring) they are clearly billed using a sample, not the population, of billed charges in the medical market.
77. Dr. Slottje states that if the population is unknown then one cannot determine if a sample is representative or not. True with the qualification of "scientific accuracy" is used as a qualification.
78. Dr. Slottje states – based on his review of the PHCS Subscriber Reference Manuals – that the Ingenix data collection effort is a carefully explained and transparent process.
79. The Subscriber Reference Manuals are summaries of the process contained in 50 or 60 pages. The Ingenix processing is described in a 600 page document and even these 600 pages do not contain a full description of the processing. Moreover, Ingenix efforts to induce data contribution in compass a range of activity which is not described in the Subscriber Reference Manuals.
80. Ingenix keeps statistics on data contributors. Also, as provided in the August 9 Report on the Merits, it is possible to provide information about who is contributing data (although contributor data information is included in a confusing array of reports) and the level of billed charges for each of the contributors. Such information and statistics are not provided to health insurers, to employers, to providers or to patients.
81. Even if information is provided, it is an open question regarding whether recipients can understand it.
82. It would be improper-as Dr. Slottje-to conclude that the Ingenix data collection effort is carefully explained, transparent or even understood by health insurers and it would be clearly erroneous to draw such a conclusion for employers, providers or patients.
83. As noted in the August 9 Report on the Merits, the information about the amount of data collected and the level of billed charges provides important information that should not be ignored in any analysis of representativeness. Dr. Slottje, however, elects to do so in his rebuttal report.
84. A few firms dominate the Ingenix data collection. Two of them, CIGNA and Aetna, contribute billed charges that are substantially lower than the rest of the contributors. Given the firms that contribute and the firms that do not, data suggest the basis for a reasonable inference that billed charges for non-contributors are higher than for the contributors.
85. Ingenix prepares and provides the percentile products. It holds them out as accurate and



representative. It is Ingenix, not providers or patients, who should bear the burden of establishing the representativeness of the contributed data.<sup>15</sup> Moreover, there are a wide range of techniques available to Ingenix to ascertain the representativeness of the contributor data such as scientific surveys. Ingenix has steadfastly refused to do so.<sup>16</sup>

#### **Modifiers**

86. Professor Slottje criticizes the discussion of modifiers on a number of grounds:
- That the Ingenix decision to exclude and include modifiers is not "ad hoc." Since Ingenix only scrubs modifiers that affect payment the report is erroneous.
  - Ingenix appropriately eliminates claims with modifiers that impact payments and appropriately retains claims with modifiers that do not impact payments.
87. Dr. Slottje's critique distorts the discussion regarding modifiers in the August 9 Report, accepts summary statements by Ingenix employees with no grounds for support and clear scientific evidence to the contrary in the August 9 Report and takes numerous deposition statements out of context.
88. Clearly, if modifiers distort billed charge the inclusion of claims with modifiers when constructing percentiles will provide distortions in the percentile data products. Ingenix excludes some claims with modifiers in constructing percentiles but includes many others. The basis for inclusion and exclusion has never been explained by Ingenix nor has any scientific investigation justifying the inclusion and exclusion been produced.
89. Dr. Slottje suggests that the Ingenix approach is not ad hoc. The common meaning of ad hoc is a solution or process designed for a specific problem or task, non-generalizable, which cannot be adapted to other purposes. A process that is "tailor-made." The Ingenix decisions regarding modifiers – to this day – do not appear to fit a particular rule or pattern. They are by Dr. Slottje's own description, tailor-made or ad hoc.
90. Dr. Slottje appears to assume that the Ingenix decisions relating to modifiers are scientific because Ms. Carla Gee stated in conversations with him that "Ingenix only incorporates charges associated with modifiers into any given PHCS releases to the extent that the modifier does not impact the build amount of the charge." Despite this, neither Dr. Slottje nor Ingenix has provided

---

<sup>15</sup> Dr. Slottje appears to imply that inability to establish representativeness or lack of representativeness means that it is appropriate for use of the percentiles to reimburse UCR. Precisely the opposite obtains: if Ingenix and that health insurers cannot establish representativeness (and therefore lack of bias) then it is scientifically inappropriate to use the percentiles for UCR.

<sup>16</sup> Indeed, Ingenix has also failed to conduct audits of contributors' data or time series analysis in order to ascertain the accuracy and honesty of the contributed data. Dr. Slottje fails to consider this in his analysis of representativeness.

the results of scientific studies relating to whether modifiers prompt physicians to alter their billed charges as the underpinnings for the Ingenix decisions to include some and exclude other modifiers. A summary untested statement that Ingenix only retains data for claims with modifiers that do not affect billed charges has no meaning.

91. Dr. Slottje identifies the 30 most common billed charges with modifiers in 2007 and 2008. He shows a table with the 11 common modifiers that Ingenix eliminates and discusses 18 common modifiers that Ingenix includes in a footnote. While implying that he has conducted some sort of scientific test, there are no results of such test provided. Apparently the discussion merely identifies common modifiers that Ingenix includes and excludes. Dr. Slottje could have but did not attempt to investigate the impact on billed charges of the 30 most common modifiers.
92. Dr. Slottje criticizes the August 9 Report evaluation of 12 CPT codes as "nonrandom" and as non-generalizable to the remaining 7000+ CPT codes. He takes statements out of context in a confusing manner.
93. As the discussion in context with Table 10 in the August 9 Report clearly states "we used Ingenix contributor data from 2006 two 2008 to conduct a *preliminary* study." The preliminary study tested the impact of all modifiers - not just a few - in order to determine whether further investigation of modifier impact would be important, not to generalize the results to all modifiers. Nor does the Report attempt to generalize this study in any way.
94. In Tables 11 and 12 of the August 9 Report we evaluated 19 of the most common modifiers which were responsible for the 95% of all modified claims (similar to Dr. Slottje's listing of the 30 most common) by considering the difference in mean billed charges-at the CPT / geozip level-with and without them. Dr. Slottje totally ignores this discussion in his criticism.
95. Indeed, Dr. Slottje's unquestioned acceptance of Ms. Gee's contention that Ingenix only keeps modifiers that *do not* impact billed charge flies directly in the face of the findings of the scientific study reported in Tables 11 and 12. Dr. Slottje notes in his footnote 61 that 18 modifiers are retained by Ingenix because (Ingenix says) that they do not impact billed charge. However, of the 18 modifiers that are retained by Ingenix at least seven of them *do in fact impact billed charges* as described (and ignored by Dr. Slottje) in the August 9 Report: for these seven modifiers the difference in billed charges with and without the modifiers is substantial and statistically significant, requiring rejection of the null hypothesis that these modifiers do not affect billed charge and refuting Ms. Gee's contention that Ingenix only includes modified claims that do not affect billed charge.
96. Dr. Slottje ridicules the conclusion of the modifier analysis in a way that suggests that he may not understand it. The scientific analysis of the difference in mean billed charges with and without

modifiers for the 19 most common modifiers that represent 95% of the modified claims shows:

- *Overall* the 19 modifiers very slightly increase mean billed charge values-and do not bias them downward.
- Individually some modifiers increase mean billed charges while others decrease them and, only three of the 19 show "no impact."

97. Accordingly, while the inclusion of modifiers in the computation of percentiles does not appear to have an overall downward bias impact, more accurate computation of percentiles suggests that most of the common modifiers had some impact on billed charge such that the best way to deal with modified claims would be to eliminate them when calculating percentile values.

#### **Derived percentile values**

98. Dr. Slottje criticizes the illustration contained in the August 9 Report discrediting the impact of deriving percentile values. The basis for the criticism is that the illustration uses only five CPT codes at the 80th percentile so that the results are not "generalizable and that the analysis used CIGNA data rather than Ingenix contributor data." To refute the findings of the illustration Dr. Slottje conducts an analysis by comparing what happened to percentile values in the PHCS product when CPT/geozip percentile values went from derived to actual and vice versa.
99. Dr. Slottje characterizes the data derivation portion of the Report as an "experiment." However, this mischaracterizes it. The analysis is merely an illustration or simulation, not an experiment.
100. To critique the illustration on the basis that it was built only on five CPT codes using CIGNA data (and one percentile) misconstrues the nature of the selection of the CPT codes, the data and the percentile. The illustration was built using five of the most common CPT codes-evaluation and management (office visit) codes at the most common percentile used for reimbursement with data from a large major contributor with millions of claim lines. The results of the simulation are statistically significant given the number of claims involved.
101. The principal conclusions contained in the Report in the section dealing with derived data are
- derived data do not constitute contemporaneous comparative billed charges for the same or similar services in the same community when percentile values are derived by combining CPT codes or geozips.
  - The process of derivation provides distortions in percentile values.
  - Ingenix uses derived data for all of MDR and for a portion of PHCS but has never provided a comparison of the relative impacts and differences (nor has it ever explained why it would be necessary to derive all of MDR).
102. Dr. Slottje's critique fails to deal at all with these conclusions, indeed, his work supports them.

103. Dr. Slottje's empirical analysis considers 3.6% of all CPT/geozips combinations from 2001 to 2008 where actual percentiles in one time period were followed by derived percentiles in the next or the reverse. This constituted 1.5 million combinations. Dr. Slottje found that in 56% to 66 % of the cases the transition from actual to derived values increased percentiles and that in the same range the transition from derived to actual decrease percentile values. From this he concludes "there is no downward bias in the PHCS has reported percentile values caused by the use of derived charges." And also that "plaintiff's experts opinions that derived charges bias the percentiles downward are not supported by the empirical evidence. There are a full range of problems with this analysis and the conclusion.
104. First of all, the Report does not in any way conclude that (and it is not my opinion that) "derived charges bias the percentiles downward. The finding contained in the report is "the analysis requires rejection of the (null) hypothesis that the derivations produced the same percentiles as the actual data and acceptance of the null hypothesis that the derivation produce processed biases percentile values." In fact, the Report shows very clearly that the derivation produced lower percentile values some of the time and higher percentile values some of the time.
105. The Slottje analysis confirms the finding of the August 9 Report, that the derivations appear to bias percentile values.
106. The Slottje analysis totally fails to consider how his illustration might be improperly influenced by selection (and once again illustrates insensitivity to the nature of the distributions involved with the billed charge data).
107. What would cause percentiles to be actual at one time and derived the following time or percentiles to be derived at one time and actual the next? Simply, the derivations would be caused by insufficient numbers of claims to provide actual values-less than nine. As noted in other portions of the Report (and this response) reporting values with very small numbers of claims (nine to 255) produce essentially "random" results.
108. The illustration contained in the August 9 Report was selected because of the CPT codes were numerous and the results would not be influenced by small numbers claims. The Slottje illustration involves all small numbers claims.
109. The Slottje illustration fails to consider dollar value of claims rather than "counts." Accordingly, Dr. Slottje's "findings" cannot show anything empirically about downward bias in the PHCS percentile values.
110. Moreover, what Dr. Slottje's illustration does tend to show (not prove) is that the derivation process does bias percentiles-the very problem identified in the August 9 Report and the very problem shown in the illustration involving CPTs 99211 through 99215.

111. Small Numbers Issues
112. Dr. Slottje's discussion of small numbers issues is confusing, distorting and thoroughly disregards science.
113. Dr. Slottje interjects his characterization of Ingenix contributor data as a "census" then converts census to population to somehow justify what Ingenix. Some statistical definitions are clear not only to statisticians but the public:
- A population is an individual or group that represents all of the members of a certain group or category of interest.<sup>17</sup>
  - A sample is a subset drawn from the larger population.<sup>18</sup>
  - In convenience sampling the researcher selects participants on the basis of proximity, ease of access and willingness to participate.<sup>19</sup>
114. Clearly, all billed charges in a particular area would represent the population of billed charges. Equally clear, a subset of all billed charges would be a sample. The Ingenix contributor data constitute a subset of all billed charges and therefore represent a sample. To characterize the Ingenix contributor data as anything else has no meaning.
115. Since the Ingenix contributor data are provided based on the willingness of health insurers to participate, it is also appropriate to call them a convenience sample.
116. The term "census" is not a statistical term. It relates to counting a population. The Ingenix contributor data are not a census.
117. Ingenix produces percentiles that are used to pay UCR. If they are to be of any value the Ingenix percentiles must have some meaning and that meaning must relate to the concept of "usual, customary and reasonable." Put another way, if Ingenix percentiles are to have any place in the adjudication of UCR they need to bear some reasonable relationship to (be representative of) the population of the billed charges for the same or similar services in the community.
118. If the Ingenix percentile values are to have any meaning it is important that they not be "random" or so replete with error as to make them essentially random.<sup>20</sup> Thus, when Ingenix publishes percentiles for a CPT and geozip those percentiles cannot be subject to substantial random error.
119. So how much of a "sample" – how many billed charge observations – do we need in order to produce percentile values that are not random. Statisticians have investigated this question, most commonly in the context of normal (Bell curve like) distributions and statistics like as the mean,

---

<sup>17</sup> T.C. Urdan, *Statistics in Plain English*, second edition, Lawrence Erlbaum Associates, Mahwah (2005) at page 1.

<sup>18</sup> *Id.*

<sup>19</sup> *Id.*

<sup>20</sup> Although, it would appear that Dr. Slottje's view (shared by Ingenix) is that Ingenix can with impunity market random values that can be used to pay physicians.

the median and proportions (percentiles). One way that this has been posed is the construction of confidence intervals for proportions.

120. The first step is to select a "confidence level" for the study. The confidence level is the probability that the statistic of interest (the percentile here) will be within the selected confidence interval for the study. Statisticians commonly use the 95th percent and the 99th percent confidence levels depending on whether they are willing to accept 5% error or 1% error. In order to be conservative we used the 95th percent standard for the August 9 Report. However, it might well be appropriate to require use of the 99% level given the importance of the use of the percentile for payment. If Ingenix percentiles are to be used to adjudicate provider reimbursement we might well want to ensure that they are 99% accurate.
121. The test for the number of observations needed relates percentiles to proportions. If we are testing the accuracy of the 80th percentile we want to be sure that 80% of the values in our distribution are less than this amount and 20% of values are greater.
122. The second step is to determine the desired confidence interval. Given that percentiles are reported in 5% steps, for example the 75th percentile, the 80th percentile and the 85th percentile, we wish to find a percentile value for the 80th percentile that is not (by accident). To be very conservative we calculated our confidence interval for the August 9<sup>th</sup> Report as +/- 5%. Thus, we would at a minimum like to be 95% certain (or 99% certain) that the Ingenix 80th percentile lies in the interval 75 to 85.
123. The number of observations we need is related to the "z score" (a standardized value for a normal distribution) relating to the confidence level, the proportion below and the proportion above the percentile (0.8 and 0.2) and the confidence interval (here 0.0249).
- The sample size (n) needed for 95% confidence (z score of 1.96) that the 80th percentile of billed charges will be in the interval from the 75th to the 85th percentile (80 percent of the values will lie below and 20 percent above +/- 5 percent – the confidence interval or CI) is calculated as
  - $n = z^2 p(1-p) / CI^2$
  - Where  $z=1.96$ ,  $p=0.8$ ,  $1-p=0.2$  and  $CI=0.049$  there will need to be  $0.615 / .024 = 256$  observations
  - 80th percentile values reported with less than 256 observations will be the 75th or 85th percentile values more than five percent of the time.
124. If we narrow the range of the confidence interval from 77.5 to 82.5 so that there is no overlap between intervals, for the 80th percentile with a 95% confidence level we would need 991 observations! The following chart shows the number of observations needed for 95% and a 99%

confidence levels with a confidence interval of +/- 2.5%.

Percentile/confidence level	95%	99%
60 <sup>th</sup>	1487	2569
70 <sup>th</sup>	1301	2247
80 <sup>th</sup>	991	1712
90 <sup>th</sup>	558	963

125. The minimum sample size for ascertaining that the Ingenix percentiles are not random, that there is a 95% chance that they are accurate within 5% was not picked at random but was determined with regard to the sample selection formulas commonly used in statistics for normal distributions. These very conservative standards require 255 observations in order to provide nominally accurate percentiles.
126. Non-normal distributions, more narrow confidence intervals and higher levels of confidence require far more observations.
127. Dr. Slottje states "a fatal flaw with dropping all CPT/geozips accommodations with less than 255 observations is that the underlying populations of many CPT/geozip combinations fewer than 255 observations. To drop them is to delete CPT/geozips combinations with valid data for no appropriate reason and is just an incorrect procedure to perform. Dr. Slottje then quotes deposition answers from a confusing series of hypothetical questions to characterize my "contentions" as "puzzling."
128. There are deep and profound problems with Dr. Slottje's views regarding these small numbers issue.
- Any reference to populations here has no meaning because Ingenix does not know what the population size for any CPT/geozips combination is or whether the contributor data has captured the population. Therefore, the Ingenix analysis must start with the premise that it is working with a sample. Indeed, the fundamental flaw with the Ingenix approach is that it assumes that it has captured all of the billed charges for a CPT/geozips combination when it has not.
  - Further, Dr. Slottje has been either unable or unwilling to perform a simple sample size calculation, either does not know or does not want to know the sample size formulas.
  - And even more fundamental, Dr. Slottje totally fails to understand the nature of percentiles and the meaning of a billed charge distribution.
  - There are 20 five percentile ranges in a data distribution. If the data were distributed "flat" and there are only 10 observations, it is impossible to allocate the observations to percentiles even if the distribution represents a population.



- Even if Ingenix captures a population of 20 observations and the distribution is flat there will be one observation per five percentile point-hardly enough to permit any comparison as UCR.
- If Ingenix captures a population of 200 observations that is essentially flat there will be 10 observations per percentile-coming close to a meaningful comparison group.
- If the population follows a normal curve distribution or one that is highly skewed right (more likely with physician billed charge data) substantially more than 200 observations will be required in order to produce enough comparative information at the higher percentiles. In order to get 10 observations at the 80th percentile in a normally distributed population we will need substantially more than 250 observations.
- Put another way, only if Dr. Slottje can identify a complete population can we determine how many comparison billed charges that will be produced at any given percentile and only then can we determine if we have sufficient data to make any UCR comparisons. Failing that, we are working with samples and statistical science that tells us how big are sample size needs to be in order to have appropriate confidence that are percentile is what it purports to be.

**Billed charge inflation**

129. The August 9<sup>th</sup> report describes the time and efforts involved in collecting data, processing data to create percentiles, distribution of percentiles and incorporation of new percentiles. The report also considers the ongoing provider billed charge inflation over time contained in the in Ingenix contributor data and in Aetna and CIGNA billed charge data indicating Ingenix, Aetna and CIGNA know or should know about it.
130. Dr. Slottje states that I "believe at inflation in billed charges may be adversely impacting the percentiles or that the UCR are determined using them." Dr. Slottje also states that there are no principles or laws in statistical science that suggests the appropriate construction of the data set requires an adjustment for inflation and that whether or not one adjusts for or measures inflation is a controversial subject and ultimately requires a value judgment. Finally, he contends that Ingenix explicitly discloses a method he uses to collected data each and every year in its pH CS reference manuals.
131. The discussion relating to billed charge inflation is not a matter of "belief" but relates to several clear facts:
- The percentiles that the health insurers use to reduce patient and provider payment based on UCR are not constructed from contemporaneous billed charge comparisons but from billed charge data collected at some time in the past, possibly, as noted in the Report, as much as two years prior.
  - At the time that Ingenix sells and distributes the percentile products, the percentiles do not



reflect current billed charges but past billed charges.

- Ingenix issues for releases each year for its MDR product and includes an inflation multiplier or adjustment in the MDR product. For the PHCS product (which Aetna and CIGNA use) Ingenix issues a new release two times each year and does not include an inflation multiplier.
- The Ingenix contributor data show billed charge inflation that can be modeled with a coefficient relating to time that is positive (about 1/2% per month) and statistically significant. Aetna and CIGNA claims data show the same trend.
- Given the fact of positive billed charge inflation the percentiles contained in the Ingenix product are lower than contemporaneous billed charge percentiles.
- That the percentiles used to deny payment are not current billed charge percentiles but percentiles constructed using data from one or two years prior may be "transparent" to health insurers if they carefully consider the Reference Manuals but they are mysteries to patients, providers and many employers.

132. Dr. Slottje states that measuring and adjusting for inflation after the fact (what he calls past inflation) is not a statistical issue and in fact requires a value judgment as to whether or not to address for inflation and, if so, how. He states that measuring and adjusting for future inflation is a complex and can't pursue problem that I "admit" that Ingenix is not obliged to do.
133. With this oversimplification Dr. Slottje mischaracterizes the entire problem in order to dismiss it. Health insurers use percentile data for UCR. UCR, limitations on billed charge payment based on usual, customary and reasonable charges for the same or similar services in the community, requires contemporaneous comparisons. It relates to comparative charges at the same time, not charges one year, two years, three years or 10 years ago.
134. To say that past inflation is not a statistical issue misses the point and is, in fact, wrong. Measuring what Dr. Slottje characterizes as "past inflation" (actually, known trends in billed charges over time) is a statistical issue. And, economists, policymakers and statisticians do not "throw up their hands" in the presence of persistent inflationary trends but use of the tools of their discipline to deal with such trends. Indeed, Ingenix clearly understands how to deal with inflationary trends as illustrated by its MDR product.
135. Forecasting is a well-known, deeply explored statistical science. It includes such concepts as moving average estimation, weighted moving average, exponential smoothing, autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), extrapolation, linear prediction, trend estimation and growth curve. Time trends can be estimated using regression analysis, both linear and nonlinear. Econometric and commonly make such estimates.

136. Yes, inflation adjustment may involve professional judgment but economists and policymakers deal with inflation all the time. That various professionals may argue about specifics does not mean that we should reject inflation adjustment because it is controversial.
137. Dr. Slottje states that I am "advocating" that Ingenix adjust for inflation in its PHCS product, that I admit that "my proposed process" is impractical or that I suggest that Ingenix adjust percentile values to account for projected future inflation. He states that I "concede" that Ingenix is under no obligation to adjust for inflation. These statements are distorted and untrue. I have not and do not "advocate" for any particular Ingenix process or product. My deposition testimony responded to a number of confusing and distorted hypotheticals relating to what I would do if I were providing a percentile data product. I do not have a "proposed process." I do not concede that Ingenix is under no obligation to adjust for inflation or that understanding and knowing about inflation, Aetna and CIGNA are under no obligation to deal with the issue. The obligations of Ingenix, Aetna and CIGNA are matters for proper determination by the court.
138. To be perfectly clear (given the deposition questions), if I were providing provider billed charge percentile data products in an era of billed charge inflation, I would take steps to make my product as accurate as possible and those steps would include adjustment for billed charge inflation that is as scientifically sound as possible. Indeed, I might well offer forecasts or adjustments that make the billed charge product contemporaneous and accurate. To do so would not be impractical or controversial. It would use science, not ignore it.
139. In the end, however, Dr. Slottje's entire discussion misses or obscures the point. The question here is not whether or how one adjusts for inflation but whether the Ingenix percentile values are biased downward. The nature of the Ingenix data collection, data processing and use of percentile values to deny patient and provider payment shows that they percentile values are, in fact, biased downward. This has nothing to do with inflation adjustment or forecasting.

**Discussion relating to a "hypothetical database"**

- On page 30 of his Response Dr. Slottje notes that I "contend" that "my proposed hypothetical database those quotes should not suffer from an alleged inflation bias. He states that "I simply assume that Aetna is required to use the 80th percentile of my hypothetical database."  
"Reference to a specific parameter for reimbursement could potentially vary between different percentile levels, the mean or median amount."
- I do not propose a hypothetical database
- I do not assume that Aetna is required to use the 80th percentile of any hypothetical database
- The discussion at page 30 appears to have no reference framework whatsoever

140. Dr. Slottje contends that the 80th *percentile* values for the majority of the CPT /geozip combinations in the PHCS database are higher than the corresponding *mean* values for the CPT/geozip combinations in the 300 CPT study of downward bias in the Ingenix PHCS product. Dr. Slottje contends that the positive impact of application of the 80th percentile could overcome overall negative influence from the flaws in the PHCS methodology.

- This analysis confuses and confounds percentiles and means as well as the particular CPT/geozip combinations overall and in the 300 CPT study.
- Nowhere in my report or testimony have I assumed or required Aetna to use the 80th percentile of any database. In fact, much of Aetna's out of network payment uses the 80th percentile as a payment limiting benchmark but the work in the report does not assume or require it.
- The 300 CPT study offers no particular insight into PHCS processes individually. It simply compares percentile values for the contributor data (data obtained from contributors prior to Ingenix processing) and the PHCS product percentiles in order to consider whether the PHCS percentiles are biased. Dr. Slottje either misunderstands or distorts it.
- The 300 CPT study involves no analysis of mean values. Neither Ingenix nor CIGNA limit out of network reimbursement based on mean values.

141. Dr. Slottje contends that "the variance in the particular percentile of the Ingenix PHCS database could, under numerous circumstances, increase reimbursement levels such that any downward bias is outweighed in the aggregate effect may be positive for plan participants." It is difficult to understand this assertion.

142. The 300 CPT study and the 350 CPT study -into the comparison of MDR to PHCS-contained in the Report as well as the 500 CPT study and the 5000 CPT study provided in this response show the relationship of contributor data percentiles compared with product percentiles. All of these studies show one thing – that the PHCS percentiles are biased downward. No amount of "variance" can overcome this.

#### **Bias in the PHCS data**

143. Dr. Slottje contends that the analysis of bias in the PHCS percentiles suffer from a number of flaws among the most serious of which are:

- use of an incorrect unit of analysis

- application of a self created filtering methodology to a smaller database that bears no resemblance to the actual Ingenix contributor database
  - makes mistake in creating percentiles
  - uses the wrong percentile data due to a copying error
144. Dr. Slottje contends that in his July 30, 2010 report, he compared contributor data percentiles to the PHCS percentiles. He did not. Dr. Slottje added back the billed charges that had been eliminated by Ingenix in its high low screen, comparing the effect of reverse engineering the high low screen.
145. Dr. Slottje contends that I have taken a "meaningless approach" to estimate series of downward bias. He states that "common sense would suggest that if one wanted to assess the impact of the actual high low scrub employed by Ingenix in a straightforward way" one would simply reverse-engineer the high low screen. Dr. Slottje's analysis assumes that the high low screen is the only problem in the Ingenix data. To the contrary, there are numerous known problems and there may well be others that are not known.
146. Dr. Slottje and Ingenix appear to take the position that the Ingenix percentiles can be anything that Ingenix wants them to be. However, given that Ingenix markets percentiles to adjudicate UCR and given that health insurers use the percentile to adjudicate UCR a more reasonable position would be that the percentiles should be as reflective of contemporaneous billed charges for the same or similar services in the community as possible.
147. The most accurate way to evaluate whether the Ingenix percentiles truly reflect UCR would be to obtain all contemporaneous billed charges for each medical care procedure in an appropriate area (the population of billed charges), calculate percentiles from the population and compare them with the Ingenix product.
148. However, as described in the Report, given limited time and resources our efforts have been limited to a reasonable second best approach: using the Ingenix contributor data (only available for 2006 through 2008) without the problem processes to compute percentiles for comparison with the Ingenix product percentiles. This is not a meaningless approach. It seeks to determine whether the Ingenix processing (all of it, not just the high low screen) provides a source of bias. Use of the contributor data does not permit direct evaluation of bias attributable to lack of representativeness but it does permit an inference about representativeness.
149. Dr. Slottje states that the studies use the wrong unit of analysis. "The appropriate unit of analysis in this matter is of course the CPT/geozip combination." Dr. Slottje is wrong. The unit of analysis applied in the 300 CPT study, the 350 CPT study, the MDR/PHCS comparison, the 500 CPT study and the 5000 CPT study all use the CPT/geozip as the unit of analysis.

150. Dr. Slottje contends that I “created my own database.” This is wrong. The Report and clarifying testimony clearly describe how the CPTs and geozips were selected for analysis and comparison.
- The 300, 350 and 500 CPT studies selected the most common CPT codes from the contributor data in order to obtain as much meaningful information as possible in the short time that was available to conduct the study.
  - It is theoretically possible, even likely, that the CPTs selected are not representative of “all” CPT/geozip combinations (particularly given that the great majority of the combinations have zero claims or insufficient numbers of claims).
  - However, the omission of the CPTs that were not studied – since the CPTs were selected to cover 80%, 90% and 95% of all claims data – were (at the time of the studies) considered to be unlikely to have a material impact on the results.
151. Dr. Slottje states that I had information for all contributor data that underlies the 500,000 CPT geozip combinations in the PHCS product yet decided to perform the analysis in a “nonrandom fashion” on just 300 non-randomly selected CPT codes.
- Dr. Slottje challenges the elimination of claims with less than 255 observations
  - He also challenges comparison of percentile values in the contributor data with prior period PHCS releases in order to adjust for the influence of time.
  - He complains that the 300 and 350 CPT analysis did not adjust for numbers of units in claim lines, using a single example to illustrate his problem.
  - Dr. Slottje further complains about the elimination of modifiers.
152. True, the 300 and 350 CPTs were selected based on the most common CPT codes in order to derive the most information. This does not mean that the studies have no information – they provide perfectly good information regarding the 300 and 350 CPT selected. The question that Dr. Slottje fails to ask is whether the results of these studies can be generalized to the rest of the Ingenix CPT/geozip percentiles. He assumes that they cannot without investigation.
153. Dr. Slottje also ignores the MDR/PHCS comparison that evaluated the difference between the two Ingenix products based on all 500,000 combinations that found a significant difference between MDR and PHCS.
154. The scientific basis for the elimination of claims for CPT/geozip combinations where there are small numbers of claims as well as the need to deal with the timing issues and the impact of billed charge inflation and Ingenix’ erroneous view about the modifier effect have been discussed above. Dr. Slottje would (as does Ingenix), ignore these issues in evaluating the accuracy of the

Ingenix products.<sup>21</sup> We elected not to do so when conducting the 300 and 350 CPT studies in order to attempt to evaluate potential downward bias from all sources.

155. The units issue is of potential concern. Rather than rely on a single example to consider the impact of it we undertook an empirical study of it using the 2008 contributor data. We found that only 4% of the claims lines had multiple units and that only a very limited subset of CPTs had multiple units.
156. In order to consider the impact of disregarding units we also compared billed charge percentiles without considering units with billed charge percentiles considering units (dividing billed charges by units). To study this we selected the 500 most common CPTs<sup>22</sup> (representing 95% of all claims) and all of the geozips used in the PHCS product (more than 200,000 combinations). We compared percentile values for these 200,000 combinations with and without dealing with units. The following table shows the summary results of the comparison conducted at the 50th, 60th, 70th, 75th, 80th, 85th, 90th and 95th percentiles. Overall, the simple average difference between the two was 1.9% (1.8% disregarding the 50th percentile which was not used in the 300 CPT study). The weighted average difference was 1.6% (1.5% disregarding the 50th percentile). That the 300 CPT study and the 350 CPT study did not account for units was immaterial in terms of effect size.

**Comparison of Percentile Values  
Difference in Percentiles Before and After Controlling for Units  
500 CPT's/421 geozips from 2008 Contributor Data**

Percentile	50	60	70	75	80	85	90	95	Overall incl 50th	Overall excl 50th
Simple Average	2.5%	2.2%	1.9%	1.8%	1.8%	1.7%	1.6%	1.6%	1.9%	1.8%
Weighted Average	2.8%	2.2%	1.7%	1.6%	1.4%	1.3%	1.1%	1.1%	1.6%	1.5%

#### The 500 CPT study

157. In order to further investigate the relationship between the contributor data and the Ingenix PHCS product, particularly in light of complaints like those of Dr. Slottje, we constructed percentiles for more than 200,000 combinations of CPTs and geozip's using the 500 most common CPT codes representing 94% of all medical and surgical claims in the contributor data and all 421 geozips.<sup>23</sup>
158. We computed the 50th, 60th, 70th, 75th, 80th, 85th, 90th and 95th percentiles using the contributor data, merged these with the Ingenix product data (taken from the Noether responsive report) in order to compare the two.

<sup>21</sup> Indeed, the 300 CPT and 350 CPT studies compared the contributor percentile values with contemporaneous PHCS product as well as prior period PHCS product. Dr. Slottje ignores the contemporaneous comparison.

<sup>22</sup> Yes, not random but encompassing most of the data.

<sup>23</sup> We identified the 500 CPTs using the 2007 contributor data.

159. We compared the percentiles from the contributor data with a "contemporaneous" release of the PHCS product as well as the two PHCS products from the prior year in order to capture the effect of time but also to disregard it (as Dr. Slottje would have us do).
160. While we did not control for units (given the negligible impact of them) we considered percentile comparisons to ways: keeping observations with fewer than 255 claims and discarding observations with fewer than 255 claims in order to consider Dr. Slottje's problems regarding small cell issues.
161. The following table illustrates the results of the comparison. The comparison compels a number of findings:
- The overall impact of excluding small claims count combinations is immaterial (less than 0.2%).
  - The impact of considering past year's product releases as well as the present year is uneven. The contributor percentiles were 18.3% higher comparing the 2006 data with the first release of 2005 and 13.9% higher comparing the 2006 data with the first release of 2006 (showing billed charge inflation and an impact) but that comparison of 2008 contributor data with the first release of 2007 shows the contributor percentiles 27.3% higher while the comparison of the 2008 contributor percentiles with the first release of 2008 shows the contributor percentiles 34.2% higher.
  - Contributor percentiles -comparing contemporaneous releases and ignoring small count issues (retaining small count cells)-were 14.4% higher in 2006, 29.2% higher in 2007 and 33.8% higher in 2008. This confirms the results of the 300 CPT and 350 CPT studies and suggests that the results of those studies – used to estimate damages in the Report – 11.2% for medical surgical procedures (and HCPCS) and 9.8% for dental – were and are extremely conservative.

**Comparison of Contributor Percentiles and PHCS Product Percentiles**  
**2006-2008 Contributor Data/Current. PHCS Product in Prior Year's PHCS Product**  
**500 Most Commonly Billed CPT Codes/All PHCS Product to Use It**

All	Ingenix	50 <sup>th</sup>	60 <sup>th</sup>	70 <sup>th</sup>	75 <sup>th</sup>	80 <sup>th</sup>	85 <sup>th</sup>	90 <sup>th</sup>	95 <sup>th</sup>
2006	2005 5	9.2%	10.2%	12.9%	14.0%	15.7%	20.1%	25.9%	38.7%
	2005 11	7.6%	8.8%	11.3%	12.2%	13.9%	17.4%	22.8%	34.7%
	2006 5	5.9%	6.6%	9.0%	10.3%	12.0%	15.0%	20.5%	31.9%
2007	2006 5	10.6%	12.5%	14.3%	16.3%	20.6%	25.4%	36.7%	64.6%
	2006 11	8.1%	10.3%	12.0%	14.4%	18.1%	22.9%	33.2%	58.7%
	2007 5	10.5%	13.8%	15.3%	19.8%	24.0%	30.7%	42.4%	71.2%
2008	2007 5	12.0%	14.9%	17.3%	19.0%	22.4%	27.3%	37.7%	68.2%
	2007 11	13.3%	16.5%	19.4%	22.2%	26.9%	33.2%	46.1%	80.5%
	2008 1	13.1%	16.8%	21.2%	24.4%	28.3%	34.6%	47.4%	88.0%



> 255									
Contrib	Ingenix	50 <sup>th</sup>	60 <sup>th</sup>	70 <sup>th</sup>	75 <sup>th</sup>	80 <sup>th</sup>	85 <sup>th</sup>	90 <sup>th</sup>	95 <sup>th</sup>
2006	2005_5	9.0%	10.1%	12.7%	13.8%	15.5%	19.8%	25.7%	38.4%
	2005_11	7.5%	8.7%	11.2%	12.0%	13.8%	17.2%	22.6%	34.4%
	2006_5	5.8%	6.4%	8.9%	10.1%	11.8%	14.8%	20.3%	31.6%
2007	2006_5	10.5%	12.3%	14.1%	16.1%	20.3%	25.2%	36.4%	64.4%
	2006_11	8.0%	10.2%	11.8%	14.2%	17.9%	22.6%	32.9%	58.4%
	2007_5	10.4%	13.7%	15.1%	19.7%	23.9%	30.6%	42.2%	71.2%
2008	2007_5	11.8%	14.7%	17.0%	18.7%	22.2%	26.9%	37.4%	67.9%
	2007_11	13.2%	16.3%	19.2%	22.0%	26.7%	32.9%	45.8%	80.4%
	2008_1	13.0%	16.6%	21.1%	24.2%	28.1%	34.3%	47.1%	87.9%

**Composite Difference Contributor Percentiles-PHCS Product Percentiles  
(Weighted by Percentile Prevalance)<sup>24</sup>**

	weight	0.005	0.008	0.012	0.006	0.652	0.07	0.237	0.012	
Contrib	Ingenix	50 <sup>th</sup>	60 <sup>th</sup>	70 <sup>th</sup>	75 <sup>th</sup>	80 <sup>th</sup>	85 <sup>th</sup>	90 <sup>th</sup>	95 <sup>th</sup>	Overall
2006	2005_5	0.0%	0.1%	0.2%	0.1%	10.2%	1.4%	6.1%	0.5%	18.6%
	2005_11	0.0%	0.1%	0.1%	0.1%	9.1%	1.2%	5.4%	0.4%	16.4%
	2006_5	0.0%	0.1%	0.1%	0.1%	7.8%	1.1%	4.9%	0.4%	14.4%
2007	2006_5	0.1%	0.1%	0.2%	0.1%	13.4%	1.8%	8.7%	0.8%	25.1%
	2006_11	0.0%	0.1%	0.1%	0.1%	11.8%	1.6%	7.9%	0.7%	22.3%
	2007_5	0.1%	0.1%	0.2%	0.1%	15.6%	2.2%	10.0%	0.9%	29.2%
2008	2007_5	0.1%	0.1%	0.2%	0.1%	14.6%	1.9%	8.9%	0.8%	26.8%
	2007_11	0.1%	0.1%	0.2%	0.1%	17.5%	2.3%	10.9%	1.0%	32.5%
	2008_1	0.1%	0.1%	0.3%	0.1%	18.5%	2.4%	11.2%	1.1%	33.8%

(overall percentage calculated using prevalence of percentiles from Aetna ACAS data)

#### 5000 CPT Study

162. In order to respond to a number of critiques relating to methodology (particularly Dr. Slottje's comments regarding use of a subset of the data) we undertook a comparison of all 5000 CPT codes contained in the Ingenix PHCS medical surgical product for all 421 geozips – 2.2 million combinations.

- This study divided claims lines with multiple units in order to deal with the units issue
- The study did not drop "small claims" cells even though it would be more proper to do so
- The study avoids any semblance of selection bias by including all CPT/geozip combinations
- The study provides past and current comparisons (even though current comparisons are particularly problematic) to avoid the controversy relating to "timing" issues.

<sup>24</sup> Taken from Aetna claims data.



163. In order to conduct the study we extracted medical and surgical claims for all CPTs used in the PHCS from contributor data from 2006 through 2008. We eliminated claims with negative or zero billed charge and allowed values, with invalid CPT codes, with invalid geozips and with modifiers. We divided the billed charge by the number of units to adjust for multiple units. We generated percentile values for all CPT/geozip combinations and compared the contributor data percentiles with percentile values for the same CPT geozip combinations for the first and second releases of the prior year's PHCS as well as first and second releases of the current year's PHCS.
164. The following table summarizes the results (by weighted average percent difference) of the comparisons. A positive value indicates that the contributor data percentile exceeds the Ingenix product percentile.

**5000 CPT Study  
Summary of Results  
Weighted Average Percent Difference  
Contributor Data and Ingenix PHCS  
2006-2008 Contributor Data**

Contrib	Ingenix	50th	60 <sup>th</sup>	70th	75th	80th	85th	90th	95th	overall
2006	2005_1	0.9%	2.7%	6.6%	8.2%	10.6%	15.4%	22.5%	36.2%	14%
2006	2005_2	-0.7%	1.3%	5.0%	6.4%	8.8%	12.9%	19.5%	32.5%	12%
2006	2006_1	-2.5%	-1.0%	2.4%	4.2%	6.7%	10.3%	17.0%	29.8%	9%
2006	2006_2	-3.8%	-2.2%	1.3%	3.0%	5.5%	9.2%	15.3%	27.6%	8%
2007	2006_1	4.6%	7.9%	11.8%	14.6%	19.8%	27.0%	40.5%	69.6%	26%
2007	2006_2	3.0%	6.3%	10.0%	12.8%	17.5%	23.8%	36.6%	63.8%	23%
2007	2007_1	-0.2%	3.3%	6.8%	9.9%	14.3%	20.4%	32.6%	57.6%	19%
2007	2007_2	0.8%	5.7%	9.0%	13.8%	18.5%	26.3%	40.0%	67.5%	24%
2008	2007_1	3.1%	7.2%	10.5%	12.7%	16.4%	23.0%	34.9%	65.6%	22%
2008	2007_2	3.2%	7.8%	11.8%	15.1%	20.2%	27.6%	41.4%	76.4%	26%
2008	2008_1	2.6%	7.3%	13.2%	16.9%	21.4%	28.7%	42.4%	83.2%	27%
2008	2008_2	0.5%	5.3%	11.1%	14.2%	19.1%	26.4%	40.1%	83.1%	25%

(overall percentage calculated using prevalence of percentiles from Aetna ACAS data)

165. The results of the 5000 CPT study also show systematic downward bias in the PHCS products. The downward bias at the 80th percentile ranges from approximately 8.5% for contemporaneous comparison in 2006 (10% for prior year's comparison) to 26% for the contemporaneous comparison in 2008 (24% for the prior year's comparison).
166. Qualitatively, the 5000 CPT study produces similar results to the 300 CPT, the 350 CPT and the 500 CPT studies:

**5000 CPT Study  
Qualitative Results  
Weighted Average Percent Difference  
Contributor Data and Ingenix PHCS  
2006-2008 Contributor Data**

Contrib	Ingenix	greater	equal	less
2006	2005 1	58.8%	14.8%	26.3%
2006	2005 2	55.9%	18.0%	26.1%
2006	2006 1	51.9%	22.5%	25.6%
2006	2006 2	43.1%	32.5%	24.4%
2007	2006 1	60.3%	14.5%	25.2%
2007	2006 2	56.7%	17.5%	25.8%
2007	2007 1	53.1%	21.5%	25.4%
2007	2007 2	45.9%	30.0%	24.1%
2008	2007 1	60.3%	15.0%	24.6%
2008	2007 2	57.4%	17.8%	24.8%
2008	2008 1	54.2%	21.6%	24.2%
2008	2008 2	46.6%	30.0%	23.4%

167. Qualitatively the percentage of contributor data percentiles that exceeded PHCS product percentiles compared to PHCS product percentiles that exceeded contributor data percentiles is relatively constant at 2:1.
168. The 5000 CPT study used units rather than claim lines.

**Damage estimation**

169. In order to estimate damages the Report used the results of the 300 CPT study to estimate the "overall" or composite downward bias in the Ingenix percentiles.
170. Dr. Slottje reiterates his insistence that selection of the 300 CPTs and the 350 CPTs were not random and cannot be used for making an overall estimate. Once again, the 300 CPT and 350 CPT studies were conducted under severe time pressure. They involved more than 1 billion claim lines per year of data for three years. Those studies were performed to generate as much information as possible about the greatest number of contributor data claims.
171. The 300 CPT and 350 CPT studies found overall downward bias comparing the contributor data percentiles and the PHCS product percentiles. These findings were confirmed by the 500 CPT study and, contrary to Dr. Slottje's suggestions, by comparing all of the CPTs and geozip in the PHCS product with the corresponding contributor data percentiles.
172. Dr. Slottje refers to his high low screen study to claim that the Ingenix product percentiles are higher than the contributor data percentiles. Again, that study was limited to the high low screen -- adding back deleted billed charges. Moreover, Dr. Slottje reports the results of his high low screen

analysis perfectly backward and ignores dollar value in the changes.

- Dr. Slottje found that adding back high and low charges that had been excluded found zero impact 85% to 88% of the time and reduced percentile values 3% to 6% of the time.
- In fact, adding back the charges increased percentile values 6% to 11% of the time.
- Thus, the high low screen – overall – produced two reductions of percentile values for every increase.
- The ratio of reductions to increases found by Dr. Slottje precisely is perfectly consistent with the qualitative findings of the 300 CPT study, the 350 CPT study, the 500 CPT study and the 5000 CPT study.

### **Dr. McCarthy's Critique**

#### **Conspiracy Arguments**

173. Dr. McCarthy contends that a conspiracy among health insurers and Ingenix "does not make economic sense." That "collecting charge data, contributing charge data and buying charge data are all entirely consistent with the independent self-interest of each other many, many alleged conspirators."<sup>25</sup> However, in other sections of his report Dr. McCarthy argues that employers and health insurers have strong interests in "holding down health care costs."<sup>26</sup>
174. These assertions are actually contradictory. Of course employers and health insurers have incentives to hold down health care costs. So do patients, employees, the public and even providers. However, this oversimplifies the problem. The provider payment setting deals with allocation of scarce resources. It is what is called "a mixed motive game." There are two sets of motives here, one "zero-sum" and the second "cooperative."
- Each party has an interest in maximizing income at the expense of the others. Providers have an incentive to charge as much for their services as the market will bear. Health insurance companies have an incentive to pay providers less. Employers have an incentive to pay less for health insurance and, thereby, to induce health insurance companies to pay providers less. These interests are "zero-sum." One party's gain occurs at the expense of one or more others.
  - At the same time, all of the parties have an interest in making the system work. Excessive exercise of self interest by any of the parties to the level that would destroy the system would be counterproductive.
175. Each of the parties, therefore, has individualistic motives and cooperative motives that they must

---

<sup>25</sup>Expert report of Thomas R. McCarthy (McCarthy Report), paragraph 10 at p. 5.

<sup>26</sup> McCarthy Report at pages 15-20.

balance.

176. To the extent that one or more of the parties can make it appear to be cooperative while they exercise individualistic motives unknown to the others, they may be able to receive over allocations of resources. Put simply, to the extent that employers and health insurers promised to pay for out of network services on the basis of billed charge limited by "usual, customary and reasonable" charges for the same or similar services in the community-but pay less than this amount-employers and health insurers "win" while providers and patients "lose."
177. Further, health insurers and employers (perhaps patients) have strong incentives to compel providers to "participate" or to become in network providers. They also have incentives as noted by Dr. McCarthy to appear to employees and patients as "cooperative." To the extent that health insurers and employers can say that they are cooperative by making out of network payments based on UCR but in fact are creating strong incentives to drive providers in network by paying much less than UCR, health insurers and employers can "win" at the expense of healthcare providers.
178. Dr. McCarthy suggests (as has Dr. Noether) that conspiracy in health insurance is economically implausible because firms would merely "compete away the profits from it." Competition requires many small buyers and sellers, homogenous products, perfect information and free entry and exit into the market. Dr. McCarthy and Dr. Noether would like to believe that there is competition in the health insurance. However:
- Health insurance is now characterized by many large health insurance firms.
  - Health insurance products are not homogenous.
  - Information in the health insurance market is not close to perfect. Providers understand patient's medical care conditions at levels that health insurers cannot comprehend. Health insurers pay providers using mechanisms and systems that providers do not and cannot understand. Patients' health status is unknown to health insurers and providers.
  - Entry into health insurance markets is not "free." In order to enter the market health insurance firms must appeal to or have credibility with employers. As a factual matter there has been almost no new entry in any substantial health insurance market in the country during the past five to 10 years.
  - Self-insurance plans, Medicare and Medicaid do not compete with traditional health insurance firms for covered lives. Self-insurance is an erosion of, or opt out from, the traditional health insurance market, not a competitor. Moreover, to the extent that health insurance firms are able to pay providers less, self insured plans will be forced to use them for benefits administration.

179. Dr. McCarthy suggests that there is no need for a conspiracy to reduce provider payment rates because health insurers are perfectly free to openly reduce what they pay. This suggestion ignores the "cooperative" component of the provider payment market. One way to meaningfully reduce provider payment without angering patients and providers is to reduce payment in a way that providers and patients do not collectively understand that payment has been reduced. A conspiracy that accomplishes this has meaning.
180. Dr. McCarthy contends that the current setting "involves insurers and self-insured employers supposedly joining a conspiracy just to be able to take money out of one pocket and put it in another."<sup>27</sup> Dr. McCarthy states that setting a low percentile for out of network provider payment means higher out-of-pocket costs for employees and lower premiums for both the employer and the employee. This is not merely moving money from one pocket to another. It may well involve moving money from the employer to employees and clearly involves moving money from providers to the employer. Again, to the extent that it can be done without the knowledge of providers or employees the employer "wins."
181. Dr. McCarthy claims that August night Report on the Merits contain "a number of important errors" include (1) comparing contributor data from one time. With the HCS product data from earlier time periods, (2) including facility claims in the contributor data, (3) comparing contributor percentiles based on both the professional and technical amounts to Ingenix percentiles based on the professional amount only, (4) building contributor distributions based on claim lines without counting units and (five) including duplicate records in the analysis.
182. Dr. McCarthy relies on a declaration by James LaPorta, a medical economics manager at Aetna that notes that Aetna pays 82.8% of out of network claims subject to Ingenix-based fee schedules at full billed charges. He states "this is not what you would expect to see if Aetna had engaged in a conspiracy to reduce out of network reimbursements."<sup>28</sup> Actually, it is.
183. First, as discussed in the August 9 Report, the average billed charge for the data contributed by Aetna to Ingenix in 2007 was \$139.75 compared with average billed charges for all contributors of \$185.97. If physicians who bill Aetna submit lower charges it would be reasonable to expect that Aetna would pay a substantial number of them at full billed charge.
184. Second, as noted in the following table, Aetna pays 65.2% of its Ingenix adjudicated claims at the 80th percentile, 23.7% of its claims at the 90th percentile and 7% of its claims at the 85th

---

<sup>27</sup>McCarthy Report at paragraph 16.

<sup>28</sup>McCarthy Report, paragraph 24 at page 11.

percentile.<sup>29</sup> On average, Aetna pays at the 82.6 percentile level. Dr. McCarthy's observation that Aetna pays 82.8% of its claims in full merely reflects the percentile choices in Aetna's out of network contracts.

**Proportion of Aetna Ingenix Adjudicated Claim Lines  
by Payment Percentiles**

Percentile	Proportion	Weight
50	0.005	0.25
60	0.008	0.48
70	0.012	0.84
75	0.006	0.45
80	0.652	52.16
85	0.070	5.95
90	0.237	21.33
95	0.012	1.14
<b>Average</b>		<b>82.6</b>

185. Third, we used 2008 ACAS data to independently investigate Aetna's payment of full billed charges. We disagreed with three of the filters applied by Mr. LaPorta<sup>30</sup> but the number of claim lines that we identify as having been subject to Ingenix adjudication are similar.
- LaPorta identified 17.7 million claims lines subject to Ingenix-based fee schedules while we identified 19.5 million.
  - Both the 17.7 million claims lines and the 19.5 million claim lines represented approximately \$3 billion in charges and \$2.6 billion for the total allowed amount.
  - LaPorta reported that 82.8% of the 17.7 million claim lines were paid at the billed charge amount, which we assume means that the billed charge was equal to or less than the allowed amount.
  - We find that 79.1% of claims were paid at billed charge — close to the LaPorta findings.
186. However, of the \$3.1 billion in charges submitted, \$1.7 billion (57%) are represented by claim lines where the billed charge was paid in full. Put another way, while 20% of claim lines had a billed charge amount greater than the allowed, around 43% of charges were represented by that 20%. Considering allowed amounts, we find that almost 31% of all billed charges are represented by claim lines where the billed charge was greater than the allowed amount.
187. We also broke this down by major coding category (surgical, radiology, pathology, E/M,

<sup>29</sup> Percentages of claims paid at given percentiles taken from 2001-2008 ACAS data as described in the August 9<sup>th</sup> Report.

<sup>30</sup> For example, Mr. LaPorta eliminated claims that included both a Medicare designation and an Ingenix adjudication designation. We believe that to the extent claims were adjudicated with Ingenix they should be included, Medicare designation or not.

medicine). if we look at count using our claim lines), we found following results:

Category	Percent paid in full
Surgical	59.6%
Radiology	76.8%
Pathology and Laboratory	87.9%
Medicine	81.9%
E/M	68.3%

188. Medicine and Pathology and Laboratory are the only categories that approach LaPorta's estimates in terms of percent of claims paid in full. The others fall substantially below LaPorta's 82.8%. We would expect each of these categories to reflect 82.8% payment in full if the adjudication process is value neutral.
189. Moreover, 23% of charges are represented by Medicine and Path and Lab when billed charges are greater than the allowed amount compared with 32% of charges for all Ingenix-based claim lines. This again means that in terms of dollar volume there were more charges associated with claim lines where the billed charge exceeded the allowed amount than the actual count of claims lines reveals.
190. As shown in the following table, we also considered the percentage of claims paid in full by Aetna for evaluation and management codes 99201 - 99215 (representing a very large number of total claims) in 2004, 2006, 2007 and 2008. If the claims adjudication and payment process were value neutral (free from unknown influence) we would expect 82.8% of claims (the LaPorta overall findings) to have been paid in full for each E & M code.

	2008	2007	2006	2004
99201	81.70%	84.2%	85.1%	86.2%
99202	80.27%	81.8%	82.8%	84.1%
99203	77.53%	77.6%	78.5%	81.0%
99204	71.18%	73.3%	76.4%	78.3%
99205	70.95%	71.0%	72.6%	75.2%
99211	78.20%	81.2%	82.3%	84.1%
99212	76.65%	78.3%	79.4%	83.3%
99213	69.37%	71.4%	74.4%	79.2%
99214	65.38%	68.1%	71.0%	74.8%
99215	70.51%	70.2%	73.9%	76.0%
	71.00%	72.9%	75.4%	79.4%

191. Once again, there is a large unexplained variation in terms of percentages of claims paid in full by CPT code. Moreover, there is a pattern in the variance. The percent of claims paid in full for more expensive procedures (99205 versus 99201 and 99215 versus 99211 for example) are much lower (approximately 8%). Moreover, in 2004 approximately 79.4% of claims were paid in full (what we might expect from the LaPorta results). However, the percent of claims paid in full dropped



4% between 2004 and 2006, 2.5% between 2006 and 2007 and 1.9% between 2007 and 2008. There is no explanation whatsoever for these declines and they are inconsistent with consistent payment at the 80th percentile.

192. In addition to considering the proportion of claims paid in full for evaluation and management services by level of service and overtime we considered the proportion of claims paid in full by dollar level of claim using Aetna's ACAS 2008 data. The following table shows the results.

**Percentage of Claims Paid in Full  
by Dollar Volume of Claim  
Aetna 2008 ACAS Data**

<b>Billed Amount</b>	<b>% Paid Full</b>
1 - 100	87.5%
101 - 500	64.8%
501 - 1000	55.7%
1001 - 2000	55.1%
2001 - 5000	42.2%
> 5001	33.2%
<b>All</b>	<b>79.2%</b>

193. Overall, 79.2% of claims were paid in full. However, smaller claims (less than \$100) were paid in full 87.5% of the time while larger claims (more than \$5000) were paid in full only 33% of the time. Indeed, for all claims greater than \$100 claims were paid in full (two thirds of the time). Once again, claims denials appear to be disproportionately large for higher claim values.
194. Dr. McCarthy also notes that during 2001 to 2008 CIGNA paid more than 92.5% of its out of network medical and surgical claims at school billed charges or by using methods other than the Ingenix databases. Analysis of the CIGNA claims shows that almost none of the out of network CIGNA claims paid using Ingenix were paid in full. In addition to Ingenix CIGNA used a number of payments methods for out of network services including rental networks (what CIGNA calls wrapper network). CIGNA's use of wrapper networks has nothing to do with whether a conspiracy has not taken place or not.
195. Dr. McCarthy contends that the provider class has not suffered antitrust injury because doctors have the ability to balance bill. This contention ignores the reality of provider billing and collection and focuses only on the relationship between providers and patients.
196. When the health insurer (under traditional insurance or on behalf of a self-insured plan) surreptitiously succeeds in underpaying for out of network medical care someone loses (the patient and or the provider) and someone wins (the health insurer and or the employer).



197. If the provider balance billed and the patient paid the bill the patient is the injured party.
198. If the provider waived the balance bill then the provider was injured. To contend that neither the patients nor the provider has been injured when the provider does not balance bill ignores economic reality. The health insurer is not a third party beneficiary of the patient. Moreover, adoption of this position would be harmful a matter of public policy because it would suggest that providers never waive balance bills – resulting in substantially greater charges for patients.
199. Dr. McCarthy also suggests that there is no antitrust injury because out of network provider payments are "generous" in comparison with Medicare. The level of Medicare payment bears no relationship to whether providers and patients have been or are being paid less than the amounts that health insurers have contractually committed. Health insurers could pay at Medicare levels if Medicare levels were specified in the contract rather than UCR.
200. In paragraph 46, Dr. McCarthy gives an example of five actions by an employer's health benefit plan to reduce out of network provider payment: (1) reduction of the Ingenix percentile used to adjudicate, (2) increase in the out of network deductible, (3) increase in the out of pocket maximum, (4) additional co-payment for out of network emergency room visits and (5) increased co-payment and coinsurance for ambulatory surgery centers. This illustration underscores the point. The five actions taken by the health benefit plan were open, disclosed to employees and patients and contractually based. They involved no surreptitious undisclosed reductions of payment in ways that were not specifically provided in the contract.
201. Dr. McCarthy notes that payers do not always use the 80th percentile. This is true. However, such discussion also misses the point. To the extent that the health insurer and the employer make it clear contractually that they will use the Ingenix fee schedule for education and specify the percentile there is no problem.
202. However, to suggest that the health insurer may unilaterally determine to use the Ingenix fee schedule when the contract calls for usual customary and reasonable payment and to arbitrarily fix any level without so providing contractually and without disclosing it provides injury to patients and providers. Indeed, if the health insurer or employer is free to pick any percentile level it chooses, the health insurer and the employer would be free to pick zero as the percentile, a result that would not only injure providers and patients that would make no sense.
203. Dr. McCarthy contends that "there are too many alleged conspirators."<sup>31</sup> Not necessarily. There are only a few large national health insurers (CIGNA, Aetna, Wellpoint, Humana, and United). They could easily conspire to reduce out of network payment to providers and patients. These large

---

<sup>31</sup> McCarthy Report at page 38 et seq. McCarthy apparently forgets that he rendered the same "too many conspirators" opinion in connection with the same defendants who apparently have conspired previously.

national insurers meet continually to discuss a number of issues. They have the motive and opportunity to conspire.

204. CIGNA, Aetna, Wellpoint and Humana all have sufficient claims data to produce and market their own percentile and adjudication system. Dr. McCarthy makes much of the fact that there are economies of scale in the development of percentile data in that Aetna chose to use Ingenix because it provided broader coverage. However, most of the Ingenix data is derived and less than 500 CPT codes constitute the great majority of all billed charges. United (the owner of Ingenix) maintains massive data claims records. Inclusion of additional data from CIGNA and Aetna adds little at the margin.
205. In 1997 United "cornered the market for data benchmarking products by acquiring both MDR and PHCS. Surprisingly, no major health insurer (such as CIGNA or Aetna) objected to the acquisition of the data benchmarking products by a competitor (United) and the major health insurers have continued to provide data and to purchase percentile data products.
206. Dr. McCarthy contends that the motives of the conspirators are different. However, the motives of health insurers and employers are, as Dr. McCarthy notes, to hold down their costs while avoiding employee displeasure. The main motives of the conspirators are congruent.
207. Dr. McCarthy believes that the existence of other reimbursement methods (rental networks, Bill negotiations, Medicare fee schedules and other fee schedules disproves the existence of a conspiracy. He contends that the conspiracy would have to control for all methods used to educate out-of-network claims in order for it to be effective. Once again, this is misplaced. If the conspiracy provided economic benefit to the conspiratorial the existence of other education methods is irrelevant. Moreover, the reduction of UCR payment by using Ingenix in many cases actually facilitated implementation of alternative systems of adjudication that paid even less. Accordingly, the existence of alternative adjudication systems may actually be evidence of the existence of the conspiracy.
208. Dr. McCarthy contends that the alleged conspirators were "transparent" about how they set their "UCR's." Dr. McCarthy, like Dr. Slottje, finds "transparency" from the fact that health insurers (for themselves and as benefit administrators) and employers understand that Ingenix is being used to fix UCR. In the first instance this knowledge was not made available to employees, patients and providers. Moreover, how Ingenix fixed its percentile levels was never transparent to employers, patients and providers.
209. Dr. McCarthy contends that coordination of benefits could not effectively be used to monitor the conspiracy.

**The Exercise of Market Power**

210. Dr. McCarthy states "the Data Market appears to have substantial scale efficiencies." This is not necessarily the case. Collection of billed charge clean lines can involve billions of lines of data, but the data can be processed using a desk top computer. It would be easy for Aetna or CIGNA to enter the data market.
211. Dr. McCarthy contends that Aetna chose to adopt PHCS in 1996 due to its more extensive and more reliable coverage. However, this may not necessarily have been the case. In 1996 MDR was owned by United and PHCS was owned by the Health Insurance Association of America, an association of health insurers of which Aetna was a member. Moreover, the "extent" and "reliability" of PHCS was as much of a question in 1996 as at present. Much of PHCS then as now was comprised of derived percentiles rather than actual data.
212. Dr. McCarthy notes that uses of the Ingenix data can switch to the CMS database using an RBRVS payment system. This is true. The fact that customers may be departing a market may be as indicative of the exercise of market power as its absence. The essence of market power involves decreased quantity supplied with increased price. One would expect market departures from price increases.
213. Dr. McCarthy claims that the Linked Market for reimbursement of out of network services is not a proper market because (1) there is no separate market for buying and selling of out of network services, (2) patients can shift between in network and out-of-network providers and (3) out of network providers are neither buyers nor sellers. These observations mischaracterize the nature of the transactions in the Linked Market. In the Linked Market providers sell medical care services to patients and are paid by health insurers acting as agent for patients. There are buyers and sellers that include providers and health insurers.

**Evidence of Suppression in the Ingenix Databases**

214. Dr. McCarthy contends that a number of analyses (including his own) find there has been no systematic downward bias in the Ingenix database. Further that, his analysis points out the computational and other errors in my August 9 Report lead to erroneous findings.

**Reliance on other experts**

215. A substantial portion of Dr. McCarthy's conclusions – that the Ingenix databases have not been suppressed – relates to non-critical work of other defendants' experts submitted during the class certification stage.
216. Dr. McCarthy relies on class certification analysis by Dr. Cantor that included Medical Fees in the

United States (PMIC), Physicians Fee Reference (Wasserman), National Dental Advisory Service Comprehensive Fee Report (Wasserman) Physician Fee & Coating Guide (MAG Mutual) and Medicare Part B Physician Supplier Procedure Summary (collectively, "comparative data"). Dr. McCarthy concludes summarily – with no support – that "Dr. Cantor's benchmark analysis does provide a reasonable methodology for investigating whether the alleged conspiracy suppressed the Ingenix databases." Dr. McCarthy makes a single reference to a statement in the marriage reports that the reference of data do not constitute the population of billed charges, data correlations for them are unknown and likely as problematic as Ingenix as "missing the point."

217. The quoted statement does not miss the point. Critical analysis of Dr. Cantor's work discloses that it is based on irrelevant comparisons, ignores data collection issues related to the materials compared and contains unsupported conclusions. In important respects Dr. Cantor's work actually supports the notion of bias in the Ingenix products – both upward and downward – with downward significantly upward.
218. Despite a range of reasons suggesting that it would be inappropriate to do so, Dr. Cantor assumed that the comparative data were reliable, comprehensive and comparable to Ingenix PHCS. They are not.
- Medicare PSPS summarizes the mean value for billed charges submitted with Medicare claims at the "carrier" (the state level) by CPT code.
    - Primarily elderly Medicare patients use far different medical care procedures than younger patients insured with private commercial health insurance. The universe of CPT codes used for Medicare differs substantially from CPT codes for private commercial health insurance.
    - Billed charges for Medicare claims are known to be substantially less than comparable billed charges for non-Medicare claims.
    - The Medicare PSPS reports billed charges at the mean rather than at percentiles. Downward bias in Ingenix is minimal at the median and increases as percentiles increase. Comparison of mean values rather than higher level percentiles that are used for most payment (the 80<sup>th</sup>, the 85<sup>th</sup> and the 90<sup>th</sup>)
    - The PSPS comparison uses the wrong unit of analysis, the wrong geographic comparison and a wrong set medical procedures.
  - Both Wasserman products report billed charge percentiles by zip code - both for its medical and surgical modules and for its dental module. Wasserman claims that it obtains information by conducting surveys of physician offices. However, physician surveys classically produce low response with problematic results.

- Data sources for the Wasserman products are unknown.
- There is no indication regarding how much data is collected to produce the product.
- Wasserman fails to report the scientific methodology, if any, behind its surveys and it does not report the results of any tests for reliability, if any have been conducted.
- MAG Mutual and PMIC both report physician fees at the national level – applying a geographic adjuster.
  - Neither of these products reports billed charges by area comparable to Ingenix. Dr. Cantor "converts" the MAG regions 25 digit zip codes and the PMIC areas to CMS regions
  - Neither of them provides percentile data-Dr. Cantor merely "guessed" at comparable percentiles. Indeed, she compares two different percentile levels the 80th and the 90th to the MAG high value (without any basis for the comparison).
  - Neither of them provides information regarding how data are collected, how much have been collected and how reliable the data are.

219. Despite all of this Dr. Cantor's findings from the comparisons are consistent with downward bias in the Ingenix products. The following table reproduces Dr. Cantor's findings as set forth in Table 14 of her Class Certification report (page 24). The MAG-80 compares the MAG Mutual "high" value to the 80th percentile of Ingenix while the MAG-90 compares the MAG Mutual high-value to the 90th percentile of Ingenix. PSPS compares Medicare mean values with the Ingenix mean for all of the PSPS "regions" as well as a subset of them. "Dental" refers to the Wasserman dental data. "PFR" relates to the Wasserman medical and surgical data. PMIC denotes the PMIC comparison data.

Dr. Cantor's Comparison Data Results

Comparison	Ingenix Less	Ambiguous	Ingenix Greater
MAG-80	54%	5%	41%
MAG-90	44%	5%	51%
PSPS-all	22%	42%	36%
PSPS-subset	22%	44%	34%
Dental	68%	2%	30%
PFR	56%	2%	42%
PMIC-All	26%	57%	17%
PMIC-part	24%	52%	24%

220. The Medicare comparison of means is not particularly relevant and compares rather than percentiles, particularly the 80th through the 90th percentile. The PMIC comparison produces more ambiguity than results although the PMIC-All comparison is consistent with the notion of

downward bias in the Ingenix PHCS product.

221. Given that there is as much authority to compare the MAG high-value to the 80th percentile as the 90th, the results of the comparisons for MAG, Wasserman's PFR, Wasserman's dental and even PMIC-All are consistent with downward bias in the Ingenix PHCS.
222. Dr. McCarthy also relies on three analyses of the PHCS by Dr. Slottje: (1) a comparison of the so-called "universe" of CPT-geozip combinations (3.9 million of them) to PFR for the 75th percentile, (2) a comparison of "unscrubbed" United health group data to Ingenix at the 80th percentile and (3) in what Dr. McCarthy calls "his merits report" Dr. Slottje compare the 2006-2008 contributor data with the 75th, 80s and 90s percentiles of the 2007-2008 releases one and two of the PHCS database.
223. Once again, Dr. McCarthy's acceptance of these investigations is uncritical and fails to understand exactly what Dr. Slottje undertook and found.
224. Dr. Slottje's PFR study findings are directly opposite to Dr. Cantor's. Dr. Cantor found PFR values higher than Ingenix product values while Dr. Slottje found them lower.
- To claim that Dr. Slottje's comparison applied to the entire universe of CPT/geozip combinations is rather odd since the PFR is limited to medical and surgical claims for which the Ingenix products provide comparisons for less than 5000 CPTs in 420 geozips - 2.2 million combinations, not 4 million. Of these 2.2 million combinations, the great majority have no values or less than (using the Ingenix standard) nine claims.<sup>32</sup>
  - Further Dr. Slottje made a biased comparison. He compared the PFR values (for 2008) to the first release of Ingenix PHCS in 2008. Ingenix contributor data establishes billed charge inflation as a fact at approximately 0.5% per month. Dr. Slottje should have compared the PFR values for 2008 to both of the first and the second PHCS releases of 2008.
  - Dr. Slottje and Dr. McCarthy assume non-critically that the PFR is representative of physician billed charges and has not been constructed using estimations based on Medicare averages.<sup>33</sup> At best Dr. Slottje's PFR analysis establishes, contrary to Dr. Cantor, that the PFR values are lower than Ingenix values "on average." Further, this analysis is limited to the 75<sup>th</sup> percentile. Such analysis does not prove lack of bias in the Ingenix data or validate the Ingenix processing problems.

---

<sup>32</sup> Dr. Slottje states that he averaged PFR values to derive geozip values. In fact, simple averaging provides substantial distortions. Weighted averages are required – but PFR does not indicate the number of claims underlying the values it contains.

<sup>33</sup> PFR claims that it obtains its values from physician surveys. PFR reports values at the zip code level, not the geozip level. To claim that PFR has obtained any meaningful billed charge information using surveys in 2.2 million CPT geozip combinations is not supportable.

225. Dr. Slottje also claims that he compared of Ingenix percentiles to "unscrubbed" United percentiles. Actually, he did not. For 20 of the "most common" CPT/geozip combinations Dr. Slottje computed the percent of observations in the United data "above" the Ingenix 80<sup>th</sup> percentile for each of the six PHCS releases in 2006-2008 rather than comparing the United percentile to the Ingenix percentile. Dr. Slottje also took three random samples 384 CPT/geozip combinations of for comparison with the first PHCS release of 2004, the first PHCS release of 2006 and the second PHCS release of 2008.<sup>34</sup>
- Dr. Slottje states that "if the unscrubbed charge data contributed by United Healthcare matched the PHCS database exactly, then precisely 20% of unscrubbed charge data contributed by United healthcare for these procedures would be higher than the PHCS 80th percentile."<sup>35</sup> Not true. If the United data has only one data point at the 80th percentile which matched PHCS then we would expect at most 19% of unscrubbed data charges to be greater – not 20%. Moreover, a grouping of billed charges at the 80th percentile (say 6% of the distribution) would lead to 16% of the billed charges became greater.
  - Dr. Slottje appears to be surprised that "the amount of charges in the of the scrubbed charge data contributed by United Healthcare having values higher than the 80th percentile of the corresponding PHCS release very both positively and negatively around 20%." This is exactly what one should expect to find. The only way to do the comparison right is to calculate the 80th percentile for United and for Ingenix.
226. Dr. Slottje draws no conclusion regarding "overall" percentile levels for United as compared to Ingenix. Even if the United percentiles were higher (which Dr. Slottje fails to establish) all that would indicate is that United physicians submit higher billed charges than Ingenix's other contributors.<sup>36</sup> It says nothing about bias or lack of bias in the Ingenix percentiles.
227. In short, Dr. Slottje's United Healthcare analysis proves nothing about bias or lack of bias in the Ingenix percentiles. For Dr. McCarthy to accept it uncritically is a mistake.
228. The third Slottje study accepted uncritically by Dr. McCarthy is the reverse engineering of the high low screen. Dr. McCarthy attributed to Dr. Slottje's merits report. However the study was initially performed for and described in the Slottje class certification report dated July 30, 2010.
- Dr. McCarthy states that Dr. Slottje excluded records with coding errors, duplicates, ineligible

<sup>34</sup> It is not certain whether the United claims are from the entire year or from the six months corresponding to the PHCS product releases.

<sup>35</sup> Slottje Class Certification report at page 18.

<sup>36</sup> A look at the representativeness of the contributors shows this directly. The average 2007 billed charge for United physicians in the contributor data is \$253 while the average billed charge for all of the physician data submitted was \$186. This may be a function of the geographic areas United serves, more complicated procedures by United physicians or it may reflect higher levels of billing by physician to participate with United.



dates of service, nonexistent CPT codes, nonexistent geozip's and procedures pertaining to medical or surgical modules thereby eliminating 47% to 57% of the 1 billion records Ingenix maintains for each release.

- Dr. Slottje then focused only on CPT geozip combinations for which a high low scrub had eliminated some records. He found that Ingenix scrubbed 6% to 7% of the eligible records.
  - Finding that replacement of the scrubbed data either reduced or had no impact on 89 to 94% of the CPT geozip combinations, Dr. McCarthy says, allowed Dr. Slottje to conclude that the hypothesis that Ingenix reports charges systematically skewed downward must be rejected.
229. The previous discussion has provided substantial detail regarding the reverse engineering study. Dr. McCarthy like Dr. Slottje attributes no meaning to the fact that even in Dr. Slottje's study reverse engineering increased percentile values twice as often as they decreased them. In fact, the reverse engineering study is consistent with, perhaps even establishes systematic downward bias in the Ingenix data.
230. Dr. McCarthy then cites with approval two studies by Dr. Noether. Dr. Noether compared percentiles for CIGNA during 2001 to 2008 with Ingenix PHCS databases for 2002 to 2009. Dr. Noether compared CIGNA data from the prior year with the Ingenix database for the current year. Dr. Noether also compared percentiles in the Ingenix 2008 PHCS database to PFR, PMIC and PPS.
231. Dr. McCarthy claims that "the results of both approaches showed that there was no systematic downward bias in the Ingenix PHCS database."
- Dr. Noether claims to have compared 4.9 million CPT code geozip combinations. However, it would be hard to see how there would be enough data in the CIGNA data bank for such a comparison when only 500,000 Ingenix CPT/geozip combinations have sufficient data to permit construction of percentiles with the Lane data points.
  - If Dr. Noether is claiming equivalence for blank cells the results are indeed questionable.
232. Moreover, as has been noted with all of these studies, at best (or at worst) what they show is that Ingenix percentiles (for the PPS and MAG) not even percentiles are sometimes higher and sometimes lower than the comparison group. Such a comparison establishes nothing whatsoever regarding bias in the Ingenix data. It establishes only that the comparison data is biased in relationship to Ingenix.

#### **The "NERA" study**

233. Dr. McCarthy then conducts his own analysis of percentiles in the contributor data compared with Ingenix product percentiles starting with a so-called 350 CPT study and following with a so-

called 300 CPT study.

234. Dr. McCarthy argues that my analysis does not match the time period of the contributor data used to create the Ingenix release. Dr. McCarthy states that he "corrects time periods."
235. The August 9 Report describes several potential problems with the Ingenix data process including the potential for downward bias produced by mass elimination of data (Dr. Slottje's reverse engineering found that 6 % to 7% of the data were eliminated using the high low screen) and the potential for downward bias on Ingenix percentile values due to delays in incorporating billed charge inflation into the Ingenix product.
236. The 350 CPT study was designed to compare "apples to apples" by matching contributor claims (and percentile generated with them) to the Ingenix product release that would have been used to adjudicate them.
- For example, claims during the first six months of 2008 would not have been adjudicated using the first product release of 2008 which occurred in May.
  - At best, claims during the first six months of 2008 would have been adjudicated using PHCS release to of 2007 (from November).
  - More likely, claims during the first half of 2008 would have been adjudicated using the first PHCS release of 2007.
237. The 350 CPT study contained in the August 9 report paired contributor results with the most likely contemporaneous PHCS product. In order to be conservative the comparison both the contemporaneous product release and the most recent prior product release. This is not a mistake. It was clearly so stated in the Report. That Dr. McCarthy elected to "align" contributor data times with Ingenix products turns a blind eye to the downward bias in the products produced by delays in incorporating billed charge inflation.
238. Dr. McCarthy also claims to have "corrected" several other "important errors." What he has done is to truncate the data using a series of value judgments to produce a caricature of a study that foreordains his results.
239. Dr. McCarthy claims there is a need to eliminate claims that do not correspond to medical and surgical modules. For example, he says, "'proc code='C.'" He states that claims for the same CPT code in "various other settings" where "proc code" equals O, N, F, I or U "may" include a technical or facility charge in addition to the physician or professional fee.
- The 350 CPT study was conducted with medical and surgical CPT codes only. Our contributor data extraction limited the extraction to the "type of service" indicator for medical and surgical procedures (TOS or Type\_Svc=1).
  - In point of fact there is no "proc code" in the contributor data dictionary as described by Dr.

McCarty. There is a "proc type" which indicates "procedure type" as identified by Ingenix for its allocation to "modules." Dr. McCarthy, apparently, eliminated all procedure types other than "c."

- The concept of UCR is well known in the health care industry. The concept of "same or similar service" applies whether the services offered in an office setting, in a facility setting or whether it relates to physician services, dental services or HCPCS.
- There is absolutely no justification for eliminating data from comparative data distribution based on some unilateral decision by Ingenix (which Dr. McCarthy has failed to evaluate) relating to whether or not the billed charge "may" or "may not" include a technical or facility charge in addition to the physician or professional fee.
- Where services are rendered in a hospital or ambulatory surgery facility, facility fees separately billed and are not billed as part of a CPT code. Accordingly, one would not expect CPT codes to improperly contain facility charges or technical charges.
- Indeed, where billings for services involved a technical component the modifier "TC" is used. Our analysis eliminated claims with "TC" modifiers.
- In fact, physician office billings include professional, technical and facility components.
- It was not error to include all procedure types when evaluating the Ingenix percentiles. To the contrary, Dr. McCarthy committed error by excluding non-'c' types.

240. Dr. McCarthy contends that I did not use Ingenix' geozip in my contributor data. "He used the first three digits of the service zip code as the geozip, which matches many of the Ingenix geozip but is not always consistent with how Ingenix builds its geozip."

- A simple Google search of the Internet quickly locates the following Internet site:  
<http://www.hmohelp.ca.gov/aboutthedmhca.org/boards/fssb/notes/050419ipp.pdf>
- That Internet site contains an Ingenix publication entitled "Ingenix benchmarking products" from April 2005.
- The document states "three-digit zips are the first three digits of the zip code."
- The subroutine generates geozips using the first three digits of zip code.

241. Echoing Dr. Slottje's contentions, Dr. McCarthy argues that I did not properly count the number of units of each CPT when estimating the percentiles from the contributor data.

- As discussed in greater detail above in connection with Dr. Slottje's report, we undertook an analysis using the 500 CPT study results to ascertain whether controlling for or disregarding units we have any material impact on estimation of percentile values.
- As noted above, that analysis found an immaterial difference of 1.5% with and without

accounting for units.

- Also as noted above, we accounted for units in the 5000 CPT study the results of which found downward bias in the Ingenix percentiles of more than 16%

242. Dr. McCarthy contends that I neglected to "remove duplicate records" from analysis of the contributor data. There is no reason to believe that there aren't duplicate records in the contributor data or that even if there were they provide a source of bias.

- Ms. Carla Gee, a senior executive at Ingenix, has stated during depositions that contributors are required to eliminate duplicate data prior to submission.
- Data contributors are required to sign certifications that there are no duplicate data in the data contributions.
- We undertook an independent analysis seeking to find duplicate records and found none in the contributor data.
- If Dr. McCarthy has "found" duplicate records we would like to understand where they came from. We also would like him to demonstrate empirically how such duplicate records bias percentile computations.
- Even if there were duplicate records there would be no a priori reason to believe that they would not be distributed in the same way as other records suggesting that their existence would have no impact on analysis.

243. Dr. McCarthy contends that "for all radiology CPT comparisons" I compared (improperly, in his opinion) contributor percentiles based on both the professional and facility claim amounts to Ingenix's percentiles.

- Radiology claims include claims with a 'tc' modifier indicate that they relate to a technical component, claims with a '26' (professional component only) modifier and claims with no modifier at all.
- As discussed in connection with Dr. Slotje's report, we removed claims with modifiers from the data when computing percentiles, including claims modified by "tc" and claims modified by '26.'
- The Ingenix product percentiles also included "tc" and "pc" designations as well as designations with no such indicators. We dropped the TC and PC designation percentiles as well for the comparison so that are percentile comparisons related to unmodified CPT codes. Such a comparison would properly compare billed charges that included both professional and facility claim amounts.
- It is unclear whether Dr. McCarthy dropped claims with place of service other than physician offices from his comparison. There is no justification for doing so. Whether medical care

services are performed in the physician's office, a hospital, ambulatory care center or elsewhere, they are the same or similar service (unless there is some other medical difference between the two). Empirically, billed charges for the same procedure in the physician office compared to hospitals and ambulatory surgery centers are less, not because the hospital or ACC physician is improperly including a facility charge, because they tend to bill at higher amounts.

- If Dr. McCarthy dropped claims with place of service other than a physician office from his comparison he has bias the percentiles downward.

244. Dr. McCarthy claims to find "dramatic changes in conclusions" as a result of correcting the "errors." Dr. McCarthy focuses on the 80th percentile to try to show that nearly all of the findings of bias in the 350 CPT study disappear.

245. Dr. McCarthy then performs the same evaluation on the 300 CPT study. He states that similar to the 350 CPT study that the 300 CPT studies incorrectly compares contributor data from one time. With Ingenix PHCS data from earlier time periods, incorrectly includes facility claims, incorrectly built contributor distributions on claim lines without counting units and incorrectly includes duplicate records. Dr. McCarthy also finds that the 300 CPT study did not reflect the fact that the Ingenix PHCS dental module corresponds to a different time. Then the medical and surgical modules and incorrectly reusing the savior of contributor percentiles to compare with different Ingenix PHCS releases.

246. What Dr. McCarthy has, in fact, done is to generate a conclusion, tailoring his methodology to the desired conclusion. His conclusion, that the corrected results support the lack of systematic downward bias in the Ingenix percentiles, flows from his assumptions.

247. Indeed, Dr. McCarthy's findings are directly contradicted by the results of the 500 and the 5000 CPT studies discussed in greater detail above.

- As previously noted, the 500 CPT study includes comparisons of contributor percentiles with both contemporaneous and previous PHCS releases and incorporates comparisons for CPT/geozip combinations regardless of the number of observations.
- The 5000 CPT study incorporates all CPT/geozip combinations rather than a subset of the data, compares contributor percentiles with both contemporaneous and previous PHCS releases, incorporates adjustments dealing with units, uses CPT/geozip comparisons even where there are few observations.

248. The 500 CPT study indicates that the downward bias in Ingenix ranges from 14 to 18% in 2006, 22% to 29% in 2007 and 27% to 34% in 2008, considerably greater, not less, than the 11.2% downward bias found in the 300 CPT study that was used to calculate damages.

249. The 5000 CPT study indicates downward bias of approximately 11% for 2006, 23% for 2007 and 25% for 2008. Again, this is substantially greater than the 11.2% difference shown in the 300 CPT study that was used to calculate damages.

#### **Dr. Joskow's Critique**

##### **Billed charge damage calculation**

250. Dr. Joskow criticizes the "billed charge" calculation of damage (billed charges less allowed amounts) on basis that it "contrasts greatly with statements in his class certification report, where he suggested a much more limited application."<sup>37</sup> This is not true.
- Paragraph 45 at page 26 of my Expert Report regarding class certification dated April 6, 2010 provides "the first method to calculate class-wide damages is to establish the difference between the billed charge and the allowed amount calculated using the flawed database."
  - My Responsive Report regarding class certification dated May 1, 2010, states that "if it is determined that billed percentiles by CPT by geozip cannot adequately describe UCR, or if there are not enough billed charge data to produce percentile values with any reasonable degree of statistical confidence there will be no sound way to establish UCR so that billed charges would be the proper basis for payment."
  - Dr. Joskow ellipses a section of the initial report and juxtaposes it with the responsive report to imply that I was limiting billed charge methods to claims with insufficient data to establish UCR.
251. Dr. Joskow states that I have not provided any justification for it using a damages model that reimburses all purported class member lines based on the billed charge. The August 9 Report clearly states that the determination regarding the appropriateness of use of Ingenix percentiles as reflecting UCR is a legal determination for the court to make. The report notes that it is for the court to decide whether Aetna failed to establish reliable standards indicating usual, customary and reasonable comparisons. In such case the court could determine that Aetna's legal obligation to pay either billed charge or UCR would require Aetna to pay billed charge having failed to establish UCR.
252. The justification for providing a damages model that reimburses based on billed charge comes from Aetna's contracts with employers agreeing to pay billed charge or UCR. It is Aetna's burden to establish that its payment limitations met the standards required for UCR in the contract. Aetna's election to use dated, flawed percentile values as a proxy for UCR could well provide a justification for the court to conclude that Aetna's failure to establish UCR required it to

---

<sup>37</sup> Responsive Expert Report of Dr. Andrew S Joskow (Joskow Report) at pages 5-8.

reimburse patients and providers based on billed charge.

253. Dr. Joskow opines that the assertion that Aetna would be required to pay billed charge is "clearly absurd on its face." For this, Dr. Joskow would interpose his views on the court and would characterize contract language as "absurd." The requirement to pay billed charge absent some other limitation is not absurd. Dr. Joskow expects that Aetna will pay his billed charges for his services. This expectation is fully reasonable. Similarly, providers and patients expect Aetna to reimburse based on billed charges when it has agreed to do so.
254. Dr. Joskow somehow distorts his reasoning into requiring that billed charge must equate to usual customary and reasonable for every claim line in the reported class. He characterizes the billed charge approach as an "assumption that the 'but-for' UCR rate would equal the billed charge." This is not the case at all. There is no assumption inherent in this damage approach and it does not rest on equating billed charge with UCR in every case. The essence of the approach is that if Aetna cannot establish UCR – if Aetna cannot provide evidence of contemporaneous comparable billed charges for the same or similar services in the market – Aetna's contracts require that it paid billed charge.
255. Dr. Joskow states that I "provide no economic basis upon which one could conclude that a provider's billed charge, no matter how high, would be the proper 'but-for' UCR rate." Dr. Joskow apparently fails to comprehend that the underpinnings of this case are legal, not economic. Dr. Joskow appears to suggest that he would invalidate any contract with economic terms that did not meet his standards.
256. It is a fundamental premise of economics that "freedom to choose" allocates resources most efficiently such that trade between parties produces optimal outcomes. This premise provides that professionals render service and receive payment for services based on their billed charges for such service. Economic consultants operate on this premise. Dr. Joskow expects his bills to be paid by Aetna based on billed charge. Aetna's legal counsel to expect that their billed charges will be paid. There is nothing absurd about this at all.

**Accurate allowed damage calculation**

257. The August 9 Report provides a second damage estimate characterized as "accurate allowed." As noted, the report was prepared under severe time, resource and knowledge constraints. The "accurate allowed" approach estimated the downward bias of the Ingenix product percentiles on an overall basis and used the estimate to adjust Aetna's allowed amounts for claims that had been limited using Ingenix products.
258. Dr. Joskow notes that this methodology did not "control" for provider training. He also criticizes



the methodology for using the geozip and for failure to consider other factors such as type of facility, type of service and patient characteristics such as the patient's age and health condition. Indeed it does not. The accurate allowed damage calculation adopts the use of Ingenix percentiles and attempts to improve them based on empirical evidence of bias in the percentile data. The approach is premised on the eventuality that the court determines that use of percentile data to determine UCR would be appropriate if the percentile data were "more accurate."

259. It is ironic that Dr. Joskow criticizes the "accurate allowed" damage approach because it adopts and extends the Ingenix approach. Dr. Joskow appears to want to have it both ways: he would reject the "accurate allowed" calculation of damages because it fails to deal with important factors, yet he would not reject the use of Ingenix percentiles to establish UCR even though it fails to deal with the same factors.
260. Dr. Joskow states that I "ignore all of these allegedly flaws" when providing the "accurate allowed" damage estimation. The August 9 Report does not "ignore" these flaws at all. The report notes that it would be possible to control for such characteristics using a model-based approach but that time and resources did not permit it at that time.
261. Dr. Joskow revisits his northern New Jersey study to argue that "accounting for various characteristics in determining appropriate 'comparable' charges might 'sometimes' result in a 'but-for' UCR rate that is lower than the actual UCR rate." He further notes that even for class member claim lines where the UCR rate would have been a higher 'but-for' use of Ingenix there is no clear uniformity in the level of alleged underpayment." Dr. Joskow concludes that "determining whether there is any injury at all and calculating the damages for each proposed class member claim line based on plaintiffs' allegations would be an individualized exercise that could not be done using the simple benchmark proposed.
262. This discussion fails to comprehend UCR, the nature of the use of Ingenix percentiles to estimate UCR and their use by Aetna to limit provider and physician payment.
- UCR involves comparison of an individual claim with a group of other claims (the more the better) in order to identify claims that are "usual, customary and reasonable."
  - In order to provide a comparison the Ingenix percentiles are developed "in the aggregate" not at the individual claim line. By their very nature they involve numbers of claims, not an individual claim.
  - If the Ingenix aggregation percentiles are accurately reflective of UCR it would be appropriate to use them as an upward limit on billed charges as provided in the contract.
  - If Ingenix aggregation percentiles are biased it will be necessary to either provide a less biased set of percentile values – or as noted above to require Aetna to pay billed charge.

- Development of less biased percentile values is an aggregate, not an individual claim line exercise.
263. Understanding the percentile bias in the aggregate permits calculation of the aggregate amount of damages, the basic approach of the August 9 Report. If Dr. Joskow is concluding that aggregate damages cannot be estimated with reasonable accuracy because claims occur at the individual level, such a conclusion would be in error.
264. If Dr. Joskow is concluding that it is impossible to allocate aggregate damages to individual claims that conclusion would also be erroneous. Having used the contributor data to compute percentile values by CPT code and geozip combinations for all 5000 of the CPT codes and geozips in the Ingenix product data (described above in connection with Dr. Slottje's report) it will be possible to compare the "new allowed" amounts based on the 5000 CPT study with the amounts allowed using Ingenix products in order to allocated damages for individual claims.

**So-called "errors" in the "accurate allowed" method**

265. Dr. Joskow contends that the Aetna damage calculation contains an "error" in that it "embeds" payment of billed charges for claims involving CPT/geozip combinations with less than 255 observations. Dr. Joskow believes that this is in "error" because he believes that it is never appropriate, indeed "absurd," to pay billed charge.
266. We have discussed in great detail in both the August 9 Report and in this declaration the scientific basis for the conclusion that at least 255 observations (in actuality, this is conservative – substantially more observations might well be required for tighter confidence intervals and more limited margins for error) are required in order to produce meaningful percentile values that do not randomly indicate values for other percentiles.
267. With fewer than 255 observations-more than 5% of the time (at the 80th percentile) calculated percentile values can be expected to fall five percentiles below or five percentiles above the calculated value, essentially making the percentiles random.
268. Even if Ingenix percentile values otherwise reflect UCR, for these "small observation" CPT/geozip combinations the Ingenix percentiles do not and cannot reflect UCR because we cannot say with any appropriate degree of confidence that they are what they are represented to be. Since the Aetna contracts require payment of billed charge or UCR and because UCR values cannot be provided with any degree of scientific accuracy for small numbers of observations, we calculated damages using the bias estimate (11.5%) to adjust percentiles which had sufficient numbers of claims to establish more accurate percentile values and we calculated damages based on billed charge for the small observation CPT/geozip combinations.

269. As described in greater detail above, what happens if Ingenix percentile values cannot be used for UCR is a question for the court to determine. If the court determines that, failing use of Ingenix, Aetna should have paid billed charge – the damage calculation methodology as contained in the report applies. If the court determines that it would be more appropriate to “adjust” using the downward bias adjustment factor – even for claims with few observations – then it would be appropriate to recalculate damages. However, such recalculation should use the 20% bias factor found in the 5000 CPT study.

**Recalculating the adjustment factors for medical procedures**

270. Dr. Joskow performed calculations indicating that the same contributor data were used to generate contributor data percentile values for 2007 and 2008 in the 300 CPT study. It is possible that 2008 contributor data was inadvertently used for 2007 in the 300 CPT study.
271. Dr. Joskow excluded comparisons for the 2007 contributor data to recalculate the downward bias amount, adjusting the bias from 11.25% to 9.56% – calculating downward bias using only two years of data. This is not an appropriate adjustment. Dr. Joskow could have calculated the percentiles for 2007 but elected not to do so.
272. Contributor data were appropriately used in the 350 CPT study, the 500 CPT study and the 5000 CPT study. A more accurate recalculation would increase the bias adjustment using the 5000 CPT study to 20%, the weighted average amount by which the contributor data percentiles exceeded the Ingenix PHCS.
273. At a minimum the recalculation should be adjusted upward – from 11.25% to 20% – using all three years of data, not adjusted downward by dropping one year's worth of data. Dr. Joskow's downward recalculation is not justified.

**Recalculating the adjustment factors for dental procedures**

274. Dr. Joskow claims that I did not understand the release times for the dental product. It is true that we have not been able to understand from the record the specific timetables for cut off for the dental product, for production of the dental products, when the dental products are released and how long it takes Aetna to “load new products.
275. Dr. Joskow assumed that since Ingenix released dental modules in January and July it would be more appropriate to compare the dental percentiles for a given year with both product releases during the same year. As a result, Dr. Joskow reduced the dental adjustment from 9.8% to 6.36%.
276. Dr. Joskow's approach assumes that Aetna used the first Ingenix dental release for 2008 throughout the first six months of 2008 and the second Ingenix dental release four 2008 during

the second six months. We doubt that this occurred.

277. If Dr. Joskow can establish that Aetna implemented the first Ingenix dental module at the beginning of January 2008 and the second in January 2008 we would agree to recompute the dental contributor data comparison using these Ingenix releases. However, we would also reestimate the dental downward bias using all of the Ingenix CPT/geozip combinations.
278. However, in an effort to be more accurate we would also extend the comparison to all CPT/geozip combinations in the dental modules much as we have been able to do in the 5000 CPT study. We have no reason to believe that the results of the dental comparison would be any different than the medical and surgical comparisons in the 5000 CPT study. Accordingly, it is as likely that the general adjustment would rise as it would be for it to diminish. Accordingly, Dr. Joskow's recalculation is premature.

**Dental claims paid using medical insurance and medical claims paid using dental insurance**

279. Dr. Joskow notes that some dental procedures are covered by medical plans and that some medical plans cover dental procedures.
280. Dr. Joskow states that the August 9 Report improperly applies the medical adjustment factor to all procedures in the medical and surgical data, including dental procedures.
281. Dr. Joskow's calculations indicate that medical damages decreased by \$2.2 million by considering the proper allocation of claims and that dental damages increased by 1.6 million, a net decrease of \$600,000.
282. This is not appropriate. Dental claims are included in medical data for a number of reasons including accidents. Medical claims are included in dental data for other reasons including dental surgery. Without more information there is no reason to conclude that the dental CPTs included in the medical and surgical claims should not be estimated using the medical / surgical downward bias and vice versa.

**Incorrect definition of class member claim lines**

283. Dr. Joskow contends that the claim line definition includes claim lines with the billed charge less than or equal to the allowed amount, claim lines that are not paid at the non-preferred benefit level, claim lines where Aetna is the secondary payor in a commercial coordination of benefits situation, claim lines that were never accepted into Aetna's ACAS system and are thus invalid, and claim lines that reflect facility charges.
284. Aetna's ACAS claims system is convoluted to say the least. Aetna officials are unable to identify were clarify many of the nuances connected with it.

285. Claim lines where billed charges are less than allowed amounts were not included as class member claim lines. Moreover, the damage calculations would be reflected at zero for these claim lines even if they had been inadvertently left in the data.
286. In accordance with information that we obtained from Aetna officials, we identified claims that were adjudicated using Ingenix – those that had a Rating System Code (X\_RATING\_CD) with an HIAA prefix, and HI prefix. We attempted to exclude claims that were not the final view of a given claim. We were not aware of ways to identify claim lines that had not been accepted into the ACAS and are thereby invalid and did not understand that claims that had not been accepted into the ACAS system would be supplied with the ACAS system data.
287. We disagree that claim lines where Aetna is a secondary payor should be excluded from the damaged claim lines – if Ingenix was used to adjudicate the claims. Coordination of benefits logic applied to these claims should be appropriate.
288. We do not understand how facility charges would have been adjudicated using Ingenix. To the extent that they were so adjudicated it would be necessary to consider the appropriate context and the amounts involved.
289. Dr. Joskow adjusts medical damage estimates downward by \$8 million and reduces dental claims by \$38 million to reflect these so-called errors. We believe that most of this reduction applies to coordination of benefits settings where Ingenix has been used and is, therefore, inappropriate. Additional understanding of the facts underlying this category of “reductions” is necessary before concluding that Dr. Joskow’s observations here are appropriate.

**Date of service**

290. Dr. Joskow “claims that it is inappropriate to include claims lines with a date of service prior to 2002 when calculating damages. Dr. Joskow reduces medical damages by \$2.5 million and dental damages by \$1 million to reflect this. We believe this to be particularly inappropriate.
291. Dr. Joskow would use Aetna’s payment delays to avoid responsibility for improper payment during 2002 and later. To the extent that Aetna underpaid subscribers during 2002 even though dates of service work earlier, Aetna should be required to reimburse appropriate amounts.
292. The timing of Aetna’s underpayment should govern damage calculation, not date of service of the underlying claim.

**The class period**

293. Dr. Joskow contends that the damage estimates are overstated because they calculate damages for time periods outside of the class as defined in the complaint. Dr. Joskow states that the earliest

- subscriber plaintiffs' class period begins on March 1, 2001. Accordingly, he recalculates damages based on a March 1, 2001 class period.
294. As described in the August 9 Report, establishment of the class period is for the court to decide. The damage calculation contained in the Report estimated that the class period would extend from 1998 through the first half of 2010, noting that the time frame would be re-estimated after appropriate determination by the court.
295. Dr. Joskow's recalculation of the class period is premature. Moreover, we noted that calculation of interest would be performed after determination of the class period. If Dr. Joskow is certain about the class period he should have calculated interest.

**Breakout of damage estimates by assignments to subscribers and providers**

296. Dr. Joskow raises issues relating to whether previous settlements in MDL-1334 limit plaintiffs' recoveries in the current litigation.
297. This too is a matter for court determination particularly since the damages involved in those settlements related to improper use of physician claims processing – a different injury than payment reduction for out of network claims.
298. The basis for Dr. Joskow's calculation of percentages of claims for patients and physicians is not explained in his report. Additional inquiry will be necessary to verify this and to ascertain the impact of prior settlements on damage calculations, if any.

**Alternative adjustment factors**

299. Dr. Joskow also recalculates damages using "alternative adjustment factors" developed by Dr. McCarthy. Dr. McCarthy's approach (the NERA approach), as noted in greater detail above, uses inappropriate time comparisons and unwarranted eliminations of claims (improper limits on place of service, type of service and others) to reduce percentile levels in the contributor data.
300. Dr. Joskow's recalculation, limiting the downward bias adjustment to 1%, is therefore equally inappropriate. Indeed, Dr. Joskow's recalculation might more appropriately reflect the 20% downward bias established by the 5000 CPT study.

**Comparison of the 300 CPT study and the 350 CPT study**

301. Dr. Joskow and Dr. McCarthy contend that since the results of the 300 CPT and 350 CPT studies are not the same they are inconsistent.
302. The two studies used different CPTs and different geozips. Both studies show downward bias. The downward bias is on and "overall" similar. Not the same, but consistent.

- 303. The 350 CPT study shows substantially greater downward bias than the 300 CPT study. In an effort to be conservative we elected to use the lower downward bias estimate in the 300 CPT study to calculate damages. We could have used the 350 CPT study amounts.
- 304. Similarly, the results of the 500 CPT study show downward bias, as do the results of the 5000 CPT study. In this sense they are consistent and provide additional verification.
- 305. Levels of downward bias differ from study to study but each study shows substantial downward bias.

**Requiring sufficient observations**

- 306. Dr. Joskow states that I "provide no scientific basis for choosing 255 occurrences as a cutoff" for accepting percentile calculations for a CPT/geozip combination. Dr. Joskow, apparently, would accept the Ingenix practices of reporting percentile values for nine observations and for deriving percentile values for hundreds of thousands of CPT/geozip combinations with less than nine observations. His observation that payment of billed charge is "absurd" requires acceptance of the premise that percentile values based on nine observations in an entire geozip gives clear evidence of usual customary and reasonable billed charges and that the derivation of billed charges is an equally proper surrogate.
- 307. The implications of this assertion have been discussed in substantial detail in connection with the review of Dr. Slottje's report above. As described in greater detail above, the necessity for requiring 255 occurrences has been derived from sample size calculations using approximations to normal distribution assumptions, 5% margin for error and a confidence interval of 0.49.
- 308. Given that the physician billed charge distributions here are distinctly not normal, but log normal, 255 occurrences would give 95% confidence that the 80th percentile would lie between the true 70th and 90th percentiles. If the distribution approximates normal 255 occurrences would give 95% confidence that the 80th percentile would lie between the true 75th percentile and the true 85th percentile. Indeed, as noted above, to be 95% certain that the calculated 80th percentile is not the 75th percentile or the 85th percentile (that it lies between the 77.5th percentile and the 82.5th percentile) would require more than 900 observations.
- 309. Several other minimum sizes have been discussed. Internal Ingenix documents suggest 50 as an appropriate minimum number of observations on the basis of "returns" to scale from bootstrapping. A recently adopted FAIR Health policy would use 40, although there is no explanation for that decision. As Dr. Slottje notes that in my class certification report I used 80 as a minimum sample size based on some preliminary work. Subsequent modifications in the August 9 Report are based on additional scientific inquiry that applies sample size formulas and



formal calculations based on log normal distributions.

310. Dr. Joskow offers a number of hypotheticals regarding very small true populations where sample sizes less than 255 might be appropriate. These hypotheticals are inapt because the true population size for all Ingenix CPT/geozip combinations are not known and it cannot be assumed that they are small.
311. Dr. Joskow contends that I have not provided any analysis to show that a UCR calculated based on an insufficient number of occurrences would necessarily be biased downward such that there would be any injury or damages resulting from Ingenix's calculation of percentile values based on an insufficient number of occurrences. This observation illustrates Dr. Joskow's misunderstanding of the issue and misses the point. The problem with generating percentile values with few occurrences is it tends to produce "random" results.
- Dr. Joskow, apparently, would be happy to have some patients over-reimbursed and others under-reimbursed so long as it all balances out in the end.
  - The problem with generating percentile values using a small number of observations is that the production of large error rates produces large numbers of winners and losers-even if it does not produce downward bias.
  - Moreover, Dr. Joskow knows as does Ingenix and Aetna, that the Ingenix percentiles are being used to "approximate" usual customary and reasonable. To the extent that the Ingenix percentiles are random they bear no reasonable relationship to usual customary and reasonable billed charges. Expediency is no excuse for randomness. And to Dr. Joskow's original observations, it is not absurd to require payment of billed charge if UCR is inherently random and therefore unsupportable.

#### **Representativeness**

312. Dr. Joskow provides a range of protestations relating to my observations regarding lack of representativeness in the Ingenix contributor data. He contends that my observations are "circular," not consistent with underlying data, imply that a new mean is "correct," that UCR rates are based on upper percentile values not the mean, and that the 12% bias estimate may reflect different insurance companies having different billed charge compositions by procedure and geographic area.
313. What Dr. Joskow would like to "will away" is the "fact" that mean billed charge values for two large contributors to the contributor data are substantially lower than the mean billed charges for the remainder of the data contributors. He cannot. There are substantial disparities in the billed charges submitted by different contributors.

314. I stated clearly in the August 9 Report that because the population of billed charges is not known, there is no way to establish with scientific accuracy whether the Ingenix contributor data are representative of physician billed charges. That means for any given CPT/geozip combination, for any aggregation of the data or for the data taken altogether there is no way to tell whether the percentile values produced bear any relationship whatsoever to the actual or real usual customary and reasonable billed charges in an area. The most telling result of that observation is that it would support a finding that Ingenix percentiles can never be used to approximate UCR.
315. The statement (not argument) that the presence of two large national health insurers whose billed charges are substantially lower than other contributors provides a factual basis for an inference of the fact of and amount of representativeness bias is not circular.
- The observation that large national insurers have bargaining power and pay physicians less is an attempt to explain why the disparity might exist and is not driven by it.
  - Dr. Joskow is right in noting that the fact that United is large but the mean value of its billed charges is higher likely contradicts the bargaining power explanation.
  - Whatever the explanation, the fact is that mean billed charge values for Aetna and CIGNA are substantially lower than mean billed charge values for the other contributors.
  - I do not suggest that Aetna and CIGNA's payment to physicians relates to billed charges. But there is a relationship that Dr. Joskow does not understand. The great majority of the data contributed by Aetna and CIGNA are for in network claims. If Aetna and CIGNA pay poorly they may "attract" physicians who are more likely to bill at lower amounts - who are willing to accept payment at lower amounts. As a result, there may be a selection effect in terms of the type of physician who will agree to network terms with Aetna and CIGNA.
  - Dr. Joskow states that I "allude" to issues regarding profiling of data by Aetna but provide no empirical analysis of the issue. My representativeness discussion in no way relates to Aetna profiling. However, Dr. Joskow may be right, that profiling may be a reason for the differences.
316. I do not at all contend that by taking Aetna and CIGNA out of the calculation that the new mean is "somehow correct." The exercise is a classic "missing data" exercise in modeling. We are missing data from health insurance company contributors who have not provided data to Ingenix. Do we have any basis for inferring what they look like?
- If we think they look like Aetna and CIGNA, then we have reason to infer that the missing data contributors would provide lower levels of billed charges.
  - If we think they look like the data contributors other than Aetna and CIGNA then there is reason to infer that the missing data contributors would provide higher levels of billed

charges.

- Because all of the large health insurers already contribute, I suggest that the missing contributors look like the mix of contributors other than Aetna and CIGNA.

317. Dr. Joskow notes that the UCR rate is based on one of the upper percentile values, not the mean. Dr. Joskow here correctly notes that the mean is more greatly affected than the median and other percentiles by extreme values or outliers.

- Even more to the point, very large data sets are not particularly influenced by outliers.
- The search for missing data relates to the entire data distribution (therefore the mean) rather than missing data for medians or any particular percentile. Use of the mean to evaluate implications about what is missing is appropriate.

318. The August 9 Report reflected as "unknown" (based on information provided by Ingenix) a CIGNA subsidiary that had high mean billed charges (672.52 nationally, 753.14 for New York).

- Correcting for this omission, mean billed charges for non-Aetna and CIGNA contributors nationally (for 2007) were \$206.55 compared with mean billed charges for all contributors including Aetna and CIGNA of \$185.97. The mean billed charges for contributors with Aetna and CIGNA removed were 11.1% higher than they were including Aetna and CIGNA.
- Correcting this omission for New York, mean billed charges for all contributors were \$176.25. Removal of Aetna and CIGNA increased mean billed charges for the contributors to \$186.55, 5.8% higher.
- The correction provides a very small change.

319. Removal of Aetna and CIGNA from the contributor data shows an increase in mean billed charges (for millions of claims) by 11.1% nationally and 5.8% for New York. If the missing health insurer billed claims look like the contributors other than Aetna and CIGNA we would estimate their billed claim charges to be 11% higher than the Ingenix contributors. To be conservative we estimated that, on average, the missing health insurers billed charges would be 5.6% higher on average.

320. Dr. Joskow speculates that the difference may reflect different insurance companies having different billed charge comparisons by procedure and geographic area.

- That is why we specifically tested New York.
- However, given that billed charges for New York are 5.8% higher while billed charges for the rest of the nation are 11.1% higher, there will be areas (within the contributor data) where contributors in states other than New York are 17% higher.
- Moreover, given the tremendous volume of contributor claims and the national scope of the contributor data it is very unlikely that the reason for the difference is different compositions

of procedures and varying geographic area.

321. Dr. Joskow states that "even if a comparison based on percentile values of billed charges at the procedure code/geozip/year level is performed and the results show that the upper percentile values of billed charges associated with Aetna and CIGNA are lower... This would not mean that any representativeness bias exists." True, but it would provide a relatively strong inference.
322. Dr. Joskow observes that the representativeness damage estimate is 50% of the accurate allowed amount. True. The downward bias estimate generated from the 300 CPT study that we used to compute damages in order to be conservative was 11.2%. The lower value New York representativeness difference estimation was approximately one half of this, so we estimated representativeness damages at one half of the downward bias damage estimate.
323. Contrary to Dr. Joskow's conclusion there is (are) a basis for the representativeness damage estimate. The bases for the representativeness damage estimate include
- the "fact" that mean billed charges for Ingenix contributors other than Aetna and CIGNA are 11.1% nationally and 5.8% for New York – higher than mean billed charges for all Ingenix contributors including Aetna and CIGNA.
  - the existence of grounds for a reasonable inference that missing contributors look more like the contributors other than Aetna and CIGNA.
  - the conclusion that if the missing contributors look like the non-Aetna/CIGNA contributors, inclusion of their billed charges in the Ingenix contributor data would increase billed charges (and percentile values) by 5.8% (New York) to 11.1%.
  - the conclusion that representativeness damages may be inferred at the level of 50% to 100% of the downward bias damages. To be conservative we estimated using 50%.

#### **Association damages**

324. Dr. Joskow criticizes the Association damage estimate as "severely flawed" rendering the damages estimate unreliable.
325. Dr. Joskow states that I have not provided a framework for the 'but-for' world in which employees of the plaintiff medical associations and societies would not have spent time dealing with out of network payment issues underlying this litigation.
- I assume that Dr. Joskow is not alleging that the various associations have lied about their use of time and resources – not to deal with this litigation but the to deal with out of network provider underpayment related to Ingenix.
  - Dr. Joskow states that I have failed to demonstrate that the Association plaintiffs would have been paid any less in the "but for" world or that they would not have engaged in "valuable

alternative activities."

- Dr. Joskow would, apparently, require as a condition for recovery that the association plaintiff employees – absent the improper conduct – would have had nothing else to do.
- Dr. Joskow's sideways allusion to opportunity cost is precisely the point. Defendants' conduct required the Association plaintiffs to use their time for pursuits that would have returned at least the value of the time of the employees to them had they not been required to deal with the out of network payment problem.

326. Dr. Joskow appears also to require that the associations "hire additional employees" as a condition for recovery. Dr. Joskow's observations are of the nature of legal conclusions and appear to be particularly inappropriate.

327. Dr. Joskow criticizes my survey as requiring respondents to "recollect the time spent" on tasks over a number of years dating back to 2004. This raises questions about the accuracy of the estimates. Something Dr. Joskow would relate to "recall bias."

- If Dr. Joskow disbelieves the association estimates it would be appropriate for him to say so.
- The estimates have been recorded for the record. The associations may be called for depositions and at trial and these recollections tested for veracity.

328. Dr. Joskow criticizes the survey for its informality. He cites survey research literature relating to survey design and execution and response verification. He cites a "generally accepted principle" that "the attorney" should have no part in carrying out the survey.

- All of these objections would be quite apt in the context of a large impersonal survey that sought to elicit complex information or opinions.
- The very reason for survey design, execution and response verification is the determination of representativeness (therefore generalizability) of responses and whether survey respondents have provided truthful and accurate answers.
- If Dr. Joskow believes that the informal nature of the survey has somehow tainted its results he should say so.
- In fact, the association plaintiffs are a finite number. They have been deposed and can be required to testify at trial. If Dr. Joskow doubts the truthfulness of the responses to the informal survey he may have legal counsel make inquiries through depositions or at trial. If the trier of fact doubts the truthfulness of the responses she is free to reject them.

329. Dr. Joskow states that I provide no indication of how I derived my salary estimates and how they apply to individual employees across all of the medical associations and societies and all of the years.

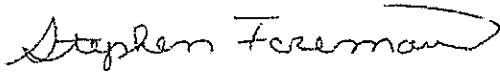
- I have spent the past 10 years of my career serving as a senior executive in a medical society

and as associate provost of a major university.

- As part of that service I have prepared program budgets, have been involved with medical society budgets and have recruited executives on a regional and national basis
- My estimates for executive salaries are, I believe, reasonable estimates for the positions involved
- Dr. Joskow may believe that the salary estimates are wrong. He could have referred to executive salary surveys to support that belief.

330. Dr. Joskow recomputes the FMA director's damages as \$270 rather than \$1500 and at \$2,776 for MAG damages rather than \$15,150. It would be appropriate to amend the damage calculation in the report to reflect these corrections even though they appear to be immaterial.

Submitted November 24, 2010

A handwritten signature in cursive script that reads "Stephen Foreman". The signature is written in dark ink and is positioned above the printed name.

Stephen Foreman, PhD, JD, MPA